

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 18, 1999, 17:23:26 ; Search time 56.23 Seconds

(without alignments)  
373.224 Million cell updates/sec

Title: US-09-004-395-2

Perfect score: 1719

Sequence: 1 MKRKASILFFLLSTVLAFAQ.....FQNFVERISEKPESSPKN 341

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

SPTREMBL\_10.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_podent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1719	100.0	341	2 P70856	P70856 borrelia bu
2	1701	99.0	344	2 051612	051612 borrelia bu
3	326	19.0	66	2 044876	044876 borrelia bu
4	131.5	7.6	243	2 006691	006691 treponema p
5	125.5	7.3	242	2 P96127	P96127 treponema p
6	121.5	7.1	242	2 083669	083669 treponema p
7	112	6.5	4152	2 092HL3	092HL3 haemophilus
8	111.5	6.5	650	10 092S55	092S55 arabis
9	110	6.4	1805	2 092102	092102 actinobacill
10	107	6.2	947	4 014789	014789 homo sapien
11	105.5	6.1	234	2 006692	006692 treponema p
12	104.5	6.1	1119	2 051228	051228 borrelia bu
13	104.5	6.1	4919	2 092HL0	092HL0 haemophilus
14	102	5.9	241	2 083670	083670 treponema p
15	101.5	5.9	613	2 045882	045882 clostridium
16	101	5.9	2166	2 051465	051465 borrelia bu
17	100.5	5.8	318	1 057922	057922 pyrococcus
18	100	5.8	701	2 092KJ9	092KJ9 helicobacte
19	99.5	5.8	401	1 029317	029317 archaeoglob
20	99.5	5.8	763	3 060013	060013 pneumocysti
21	99	5.8	663	5 024967	024967 glarida lam
22	98.5	5.7	1927	2 054875	054875 streptococc
23	98.5	5.7	631	10 022664	022664 splinacia ol
24	98	5.7	635	3 006408	006408 saccharomyc
25	98	5.7	666	5 027147	027147 eu
26	97.5	5.7	327	2 057028	057028 bacillus sp
27	97.5	5.6	425	2 025142	025142 helicobacte
28	97	5.6	1136	3 043044	043044 schistosach
29	97	5.6	3848	5 076737	076737 dictyostell

30	97	5.6	564	9 080041	080041 bacterioph
31	96.5	5.6	336	2 045471	045471 bacillus sp
32	96.5	5.6	327	2 057539	057539 bacillus sp
33	96.5	5.6	3061	12 085265	085265 potato viru
34	95.5	5.6	815	2 047722	047722 enterococcu
35	95.5	5.6	647	10 041374	041374 splinacia ol
36	95.5	5.6	634	12 006123	006123 potato viru
37	95	5.5	1854	2 033760	033760 streptococc
38	95	5.5	570	5 025968	025968 plasmodium
39	95	5.5	647	10 041027	041027 piumm saliv
40	94.5	5.5	625	2 030704	030704 mycoplasma
41	94.5	5.5	436	2 092KD7	092KD7 helicobacte
42	94.5	5.5	649	10 040693	040693 oryza sativ
43	94	5.5	430	2 051690	051690 borrelia bu
44	94	5.5	1461	2 053070	053070 lactococcus
45	94	5.5	1522	2 0921D9	0921D9 actinobacill

## ALIGNMENTS

RESULT 1  
ID P70856 PRELIMINARY; PRT; 341 AA.  
AC P70856;  
DT 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE FLA PROTEIN.  
GN FLA.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-212;  
RX MEDLINE; 97144545.  
RA GE Y., CHARON N.;  
RT "An unexpected flaA homolog is present and expressed in Borrelia burgdorferi."  
RL J. Bacteriol. 179:552-556(1997).  
DR EMBL; U62900; AAC44770.1; --  
SQ SEQUENCE 341 AA; 38367 MW; 5F08A334 CRC32;

Query Match	Score	Length	DB 2	Length	341
Best Local Similarity	100.0%	Pred. No. 7.8e-111			
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY 1 MKRKASILFFLLSTVLAFAQETDGLAEGSKRAEPGEVLDFAEIARDPSSRLDLTYVD 60					
1 MKRKASILFFLLSTVLAFAQETDGLAEGSKRAEPGEVLDFAEIARDPSSRLDLTYVD 60					
1 MKRKASILFFLLSTVLAFAQETDGLAEGSKRAEPGEVLDFAEIARDPSSRLDLTYVD 60					
QY 1 YVYGAGGIYKPEMDVVDLGINMSVLLTPSARLQAVYKNSVAVPAVYKSESKRYAGDTI 120					
1 YVYGAGGIYKPEMDVVDLGINMSVLLTPSARLQAVYKNSVAVPAVYKSESKRYAGDTI 120					
1 YVYGAGGIYKPEMDVVDLGINMSVLLTPSARLQAVYKNSVAVPAVYKSESKRYAGDTI 120					
QY 1 LGVAVLPSPYSSQSSAMTPPKIPFYSGESGNOFLGLDINTMKEIKSVYSIGYEI 180					
1 LGVAVLPSPYSSQSSAMTPPKIPFYSGESGNOFLGLDINTMKEIKSVYSIGYEI 180					
1 LGVAVLPSPYSSQSSAMTPPKIPFYSGESGNOFLGLDINTMKEIKSVYSIGYEI 180					
QY 1 DLEVLFDPMNGMEYAYSGTLKFGWADLWSNPNYIPNISRIKIDVPYPLASSKM 240					
1 DLEVLFDPMNGMEYAYSGTLKFGWADLWSNPNYIPNISRIKIDVPYPLASSKM 240					
1 DLEVLFDPMNGMEYAYSGTLKFGWADLWSNPNYIPNISRIKIDVPYPLASSKM 240					
QY 1 FKAERVSKSHSKYKNTIFYKDLRVLYDKLVSIDSIDISESYFKYETSGTSLRLK 300					
1 FKAERVSKSHSKYKNTIFYKDLRVLYDKLVSIDSIDISESYFKYETSGTSLRLK 300					
1 FKAERVSKSHSKYKNTIFYKDLRVLYDKLVSIDSIDISESYFKYETSGTSLRLK 300					
QY 1 AHETFKRYLKLREKISTAEGSFQNFVERISEKPESSPKN 341					
1 AHETFKRYLKLREKISTAEGSFQNFVERISEKPESSPKN 341					
1 AHETFKRYLKLREKISTAEGSFQNFVERISEKPESSPKN 341					

RESULT 2  
051612 PRELIMINARY: PRT: 344 AA.  
AC 051612:  
DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN (FLAA).  
GN BB0668.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 35210 / B31;  
RX MEDLINE: 98065943.  
RA FRASER C.M., CASJENS S., HUANG M.M., SUTTON G.G., CLAYTON R.A.,  
RA LAHIGA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,  
RA DOUGHERTY B., TOMH J.-F., FLEISCHMANN R.D., RICHARDSON D., HANSON M.,  
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S.,  
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN C.,  
RA UTTERBACK T., MATTHEY L., MCDONALD L., ARTICH P., BOWMAN C.,  
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,  
RA SMITH H.O., VENTER J.C.;  
RT "Genomic sequence of a Lyme disease spirochete, Borrelia  
burgdorferi."  
RL Nature 390:580-586(1997).  
DR EMBL: AE001168; AAC67025.1; -.  
DR TIGR: BB0668; -.  
KW Flagella.  
SQ SEQUENCE 344 AA; 38834 MW; 6CD0CC1 CRC32;

Query Match 99.0%; Score 1701; DB 2; Length 344;  
Best Local Similarity 98.8%; Pred. No. 1.4e-109;  
Matches 337; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRAKSIFFLLSTVLEFAQETDGAEGSKRAEPGELVDPFALADPSTRLDLTNYVD 60  
DB 4 MKRAKSIFFLLSTVLEFAQETDGAEGSKRAEPGELVDPFALADPSTRLDLTNYVD 63  
QY 61 VYSGASGIVKPEDMVVDGINNWSVLLTPSARLAQVYVNSVAPVYSESKRRVADPT 120  
DB 64 VYSGASGIVKPEDMVVDGINNWSVLLTPSARLAQVYVNSVAPVYSESKRRVADPT 123  
QY 121 LCVRLFPSYSSOSAMIMPFRKIPFYSGESGNOFLGKGLIDINKTKELKIVSYSLGYEI 180  
DB 124 LCVRLFPSYSSOSAMIMPFRKIPFYSGESGNOFLGKGLIDINKTKELKIVSYSLGYEI 183  
QY 181 DLEVLFDNMGMEYAVSMGTLKFKGNADLIWNSPNTIPNISRRIKDDVPNTPLASSKMR 240  
DB 184 DLEVLFDNMGMEYAVSMGTLKFKGNADLIWNSPNTIPNISRRIKDDVPNTPLASSKMR 243  
QY 241 FAFRFRSKSHSKVKNFITYVVDRLVLYDKLSVSDSDSDSVFVYETSGESIRK 300  
DB 244 FAFRFRSKSHSKVKNFITYVVDRLVLYDKLSVSDSDSDSVFVYETSGESIRK 303  
QY 301 AHETFEKRVLTKREKISIAEGSFONFEKIESEKPESSPKN 341  
DB 304 AHETFEKRVLTKREKISIAEGSFONFEKIESEKPESSPKN 344  
RESULT 3  
044876 PRELIMINARY: PRT: 66 AA.  
AC 044876:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 07, Last annotation update)  
DE CHEA GENE (FRAGMENT).  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
RN  
RP SEQUENCE FROM N.A.

RC STRAIN-212;  
RX MEDLINE: 98438936.  
RA OLD I.G., TRUEBA G.A., SAINT GIRON S.I., JOHNSON R.C.;  
RT "A chek chew operon in Borrelia burgdorferi, the agent of Lyme  
disease."  
RL Res. Microbiol. 148:191-200(1997).  
DR EMBL: X91907; CAA63001.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 66 AA; 7577 MW; D6137D6 CRC32;

Query Match 19.0%; Score 326; DB 2; Length 66;  
Best Local Similarity 100.0%; Pred. No. 4e-16;  
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 DSDIDSESVFYETSGESIRKLAHETFKRVKLRKISIAEGSFONFEKIESEKPE 335  
DB 1 DSDIDSESVFYETSGESIRKLAHETFKRVKLRKISIAEGSFONFEKIESEKPE 60  
QY 336 ESSPKN 341  
DB 61 ESSPKN 66

RESULT 4  
006691 PRELIMINARY: PRT: 243 AA.  
AC 006691:  
DT 01-JUL-1997 (TREMBLrel. 04, Created)  
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)  
DE FLAA HOMOLOG-1.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
RN  
RP SEQUENCE FROM N.A.  
RA PORCELLA S.F., RADOLE J.D., NORGARD M.V.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U97463; AAB63367.1; -.  
SQ SEQUENCE 243 AA; 27299 MW; 7CB7C276 CRC32;

Query Match 7.6%; Score 131.5; DB 2; Length 243;  
Best Local Similarity 22.5%; Pred. No. 0.057;  
Matches 56; Conservative 33; Mismatches 109; Indels 51; Gaps 8;

QY 48 PSTRLDLTNYVDY---YSAGSGIVKPEDMVVDGINW---SVLLTPSARLAQVYKN 100  
DB 21 PLAAQRSKVNTQAYFIDDFDGAS-----EDGGL-AMRAAGSKFTTKGPPIIKTYEG 70  
QY 101 SVVAPAVVKSSEKRRYAGDTILGVRLVLPFSYSSOSAMIMP-----PKIPFYSGESGNOFL 155  
DB 71 MPQAVRMAAGSMQGRKKEKRFYGVCEKFNROGNNMIDLPTFGGSYELPLAGVYSG----- 127  
QY 156 GGLIDINKTKELKIVSYSLGYEIDLEVLFDNMGMEYAVSMGTLKFKGNADLIWNSPN 215  
DB 127 -----FDVWMAAGYQYRSLALVADCTGVRHTLLIGNLDQGMKNLSVSP 173  
QY 216 YIPNISRRIKDDVPNTPLASSKMRFAFRVSKSHSKVKNFITYVVDRLVLYDKLSVSI 275  
DB 174 HIPOTS-----RYLGSQHLSTFVGFRIRTPSESRVDDFYVDFDQKAL---ANNHI 221  
QY 276 DSDIDSESV 284  
DB 222 DFTIDGHEL 230

RESULT 5  
P96127 PRELIMINARY: PRT: 242 AA.  
AC P96127:  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)









ID 092HL0 PRELIMINARY: PRT: 4919 AA.  
AC 092HL0:  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)  
DE LARGE SUPERNATANT PROTEIN 2.  
GN LSP2.  
OS Haemophilus ducreyi.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-35000;  
RA MEDLINE: 99030326.  
RX WARD C.K., LUMBLEY S.R., LATIMER J.L., COPE L.D., HANSEN E.J.;  
RT "Haemophilus ducreyi secretes a filamentous hemagglutinin-like  
RT protein.";  
RL J. Bacteriol. 180:6013-6022(1998).  
DR EMBL: AF057696; AAC79761.1; -;  
SQ SEQUENCE 4919 AA; 542595 MW; 25CD44DA CRC32;

Query Match 6.1%; Score 104.5; DB 2; Length 4919;  
Best Local Similarity 19.1%; Pred. No. 2.8e+02;  
Matches 82; Conservative 58; Mismatches 119; Indels 171; Gaps 18;

OY 54 DLTNYVDVYSGAGIVKPEDWVDLGI-----NMSVLTPSARLOAY-VKNSVAVPAV 108  
DB 1001 NISNKAIVLHSLGINSLSNKKDOYVNGELIYAGNINISV-----KAHOLKNDVYKLMGI 1052  
OY 109 KESK-----RYAGDTILGVRVLFPSYSSSAMIMPF----- 142  
DB 1053 TTKTEGQASYKLYQANGHFGNDSSG-----YSEGDLNKGKFPADLNKLTVQRI 1105  
OY 142 -----KPFYSGESG--NOFLGKGLDINIK-----TMKEIKYSVYSLGE 179  
DB 1106 GRIYAGRDITFKNKSNGAGSELIIRGTI--NVNKLISYSDVSFENNMQOKVDLYTKIFE 1164  
OY 180 I--DLEVLFE-----DMNGMEYAYSMTLKEFGWADLI----- 211  
DB 1165 AKSDIELTFTKNGTHRYVLYNFKSNNEKKYRNSENKTKRSIGDLNEALSAPAEIRA 1224  
OY 211 -----MSNPYIPNISRIT--KDDVPNYP-LASSKMF 241  
DB 1225 YSSGSSSNINPVSYLALGANNSSNPYLTALTKHILGNGWQDCLKQENIKYVKQKW 1284  
OY 242 KAFRVSKSSSKYKNFIYVKDLRVLYXKLSVI-----DSDDISSEVFRVY-- 289  
DB 1285 EDFKDKGASKML-----DLYPTDKKAKIFAGIIRNGNDTISDVESEDFKKRYSK 1336  
OY 289 -----ETSGTESLRLKAHETFRVYLKREKISIAEG-----SFONFV 326  
DB 1337 FONGEWAUKNDTGTDSYDSTRKASKYKVENVDHKEKIDHKLNIKGHEITVPGVSENLN 1396  
OY 327 EKIESKEPEE 336  
DB 1397 NKNMDHQDPK 1406

RESULT 14  
083670 PRELIMINARY: PRT: 241 AA.  
AC 083670:  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)  
DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN (FLAA-2).  
GN TP0664.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 96332770.

RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.C.,  
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,  
RA SODERGERN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,  
RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTERBACK T.,  
RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,  
RA HATCH B., HORST K., ROBERTS K., MATTHEY L., WEIDMAN J., SMITH H.O.,  
RA VENTER J.C.;  
RT "Complete genome sequence of Treponema pallidum, the syphilis  
RT spirochete.";  
RL Science 281:375-388(1998).  
RN [2]

RP SEQUENCE FROM N.A.  
RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.C.,  
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,  
RA SODERGERN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,  
RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTERBACK T.,  
RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,  
RA HATCH B., HORST K., ROBERTS K., MATTHEY L., WEIDMAN J., SMITH H.O.,  
RA VENTER J.C.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE001240; AAC65637.1; -;  
KW Flagella.

Query Match 5.9%; Score 102; DB 2; Length 241;  
Best Local Similarity 27.8%; Pred. No. 6;  
Matches 42; Conservative 17; Mismatches 72; Indels 20; Gaps 7;

OY 142 KIPFYSGESGNOFLGKGLDINIKM-----KEIKVSYSLGIELDLEVLFEPMNGMYA 195  
DB 99 KVSFYR-RGYNF-----EVSSYKPLPEGVAKTVSWVWAGRGPHSLJLLEDFWGORFE 153  
OY 196 YSMGTLKFGWADLI--IWENPNYIPNISRITKD--DVPYPLASSMKRKAFFVSKSHSK 253  
DB 154 LHMGLDPSGKMLKSVAILPQHVDEKTIYOKSRFPN-----QSGNLIVGRVDCDPLEA 209  
OY 254 VKNFIYVYKDLRV--LYDKLSVISIDSDIDS 281  
DB 210 YGNIVYVFDLRLVYVDLVESREPDWSDA 240

RESULT 15  
045882 PRELIMINARY: PRT: 613 AA.  
AC 045882:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)  
DE CBM71 MOSQUITOCIDAL TOXIN.  
GN CBM71.  
OS Clostridium bifermentans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CH18;  
RX MEDLINE: 96236023.  
RA BARLOY F., DELECLUSE A., NICOLAS L., LEGADET M.M.;  
RT "Cloning and expression of the first anaerobic toxin gene from  
RT Clostridium bifermentans subsp. malaysia, encoding a new  
RT mosquitoicidal protein with homologies to Bacillus thuringiensis  
RT delta-endotoxins.";  
RL J. Bacteriol. 178:3099-3105(1996).  
DR EMBL: X94146; CAA63860.1; -;  
DR PFWA; PF00555; endotoxin; 1.  
SQ SEQUENCE 613 AA; 71172 MW; CF415276 CRC32;

Query Match 5.9%; Score 101.5; DB 2; Length 613;  
Best Local Similarity 18.8%; Pred. No. 24;  
Matches 54; Conservative 59; Mismatches 97; Indels 77; Gaps 12;

```

QY 42 AELARDPSTRLDLNIVDYVYSGASGIVKPEDMAYVDGINNSVYLTPSARLAAYANS 101
Db 208 SNLKLDPNNSWIDITRYCRFM-----TFYILDM-----ISICPIYDTKYDK-- 250
QY 102 VVAPAVVSESKRYAGDTILGVRVLP-----SYSOSSAMIMPEFKIPFYSGESGNOFL 155
Db 250 ---PINMOTLTRKYSDPVNFIDENIPISEYEKMINISPELSTLFSTLSEFTYTKSGNKF 306
QY 156 G-----KGLIDNITKMEIRVSVSLGVEIDLEVLFEPMNGMEYAY-----SMG 199
Db 307 NGHVNRHVGTDLNTNGLRETHYGNYSVEYE-SMAFDDIKAYSNNYFNNTQNNPTSVK 365
QY 200 TLKF---KGADLIWSNP-----YIPNISRIKDDVPNTPLASS-RMR--- 241
Db 366 SIKFLITKNDEWIYGEPPDSSNIDFTRNIOGYLSMLNNESSYTHSLSDMILANNDKIQINI 425
QY 241 -----FKAFRVSKSHSKVKNFIYVKDLRVL---YDKLSV 273
Db 426 DTPHSYSYSWIYKGIEDTINYISDKLINOIPLVKEVKLKSRIHSETSV 472

```

Search completed: August 18, 1999, 17:39:28  
 Job time: 962 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 16:19:14 ; Search time 543.6 Seconds

(without alignments)  
9729.313 Million cell updates/sec

Title: US-09-004-395-1

Perfect score: 1663

Sequence: 1 atgataatcttttcca...

Scoring table: IDENTITY\_NUC

Searched: 679419 segs, 1590154680 residues

Database :

GenBank: 1: gb\_ba1: 2: gb\_ba2: 3: gb\_om: 4: gb\_ov: 5: gb\_pat: 6: gb\_ph: 7: gb\_pl1: 8: gb\_pl2: 9: gb\_pl3: 10: gb\_pl4: 11: gb\_pl5: 12: gb\_ro: 13: gb\_st: 14: gb\_sts: 15: gb\_sy: 16: gb\_un: 17: gb\_v1: 18: em\_fun: 19: em\_hg: 20: em\_hum1: 21: em\_hum2: 22: em\_in: 23: em\_om: 24: em\_or: 25: em\_ov: 26: em\_pat: 27: em\_ph: 28: em\_pl: 29: em\_ro: 30: em\_sts: 31: em\_sy: 32: em\_un: 33: em\_v1: 34: gb\_hg1: 35: gb\_hg2: 36: gb\_in1: 37: gb\_in2: 38: em\_ba1: 39: em\_ba2: 40: em\_hum3: 41: em\_hum4: 42: gb\_pl4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1637	98.4	1655	2	BB062900	U62900 Borrelia bu

2	1210.6	72.8	13371	2	AE001168	AE001168 Borrelia
3	435.8	26.2	10016	2	AE001167	AE001167 Borrelia
4	314	18.9	332	1	BB062900	BB062900 Borrelia
5	164.4	9.9	2700	2	BB028962	BB028962 Borrelia
6	108.6	6.5	1260	1	THFLA1	THFLA1 Borrelia
7	108.6	6.5	1140	5	A24079	A24079 T. hyodysent
8	89.4	5.4	17356	42	AC006986	AC006986 Homo sapi
9	85.2	5.1	110000	34	PFMAL4P1_2	PFMAL4P1_2 Homo sapi
10	84.8	5.1	1128	36	MPU15677	MPU15677 Myrmecia pi
11	78.8	4.7	25307	36	PFMAL3P7	PFMAL3P7 Homo sapi
12	77	4.7	19124	35	PFVARIA	PFVARIA Homo sapi
13	77	4.6	256172	35	AC005139	AC005139 Plasmodiu
14	76.4	4.6	713	37	AF034389	AF034389 Plasmodiu
15	76.2	4.6	14667	37	AE001398	AE001398 Plasmodiu
16	75.8	4.6	108908	36	PFMAL3P8	PFMAL3P8 Plasmodiu
17	75.6	4.5	12029	37	AE001431	AE001431 Plasmodiu
18	75.2	4.5	86829	36	PFMAL3P5	PFMAL3P5 Plasmodiu
19	75	4.5	1156	36	MPU15678	MPU15678 Myrmecia pi
20	74.8	4.5	222512	42	AC006535	AC006535 Homo sapi
21	74	4.4	175374	34	PFMAL4P2	PFMAL4P2 Homo sapi
22	74	4.4	261670	34	PFMAL4P2	PFMAL4P2 Homo sapi
23	74	4.4	14001	36	PFMAL4P2	PFMAL4P2 Homo sapi
24	74	4.4	9000	36	TETRRRNA	TETRRRNA Homo sapi
25	72.6	4.4	4601	36	DMU11584	DMU11584 Drosophila
26	72.6	4.4	19517	37	DMU37541	DMU37541 Drosophila
27	71.6	4.3	3061	36	PFAROP1	PFAROP1 Plasmodiu
28	71.6	4.3	113899	36	PFMAL3P4	PFMAL3P4 Plasmodiu
29	71.6	4.3	13859	37	AE001366	AE001366 Plasmodiu
30	71.6	4.3	14135	37	AE001388	AE001388 Plasmodiu
31	71.4	4.3	2654	36	DDIDG17A	DDIDG17A Plasmodiu
32	71.2	4.3	300172	35	AC005308	AC005308 Plasmodiu
33	70.8	4.3	2426	7	SDU49822	SDU49822 Saccharomyc
34	70.8	4.3	65691	36	PFMAL3P1	PFMAL3P1 Plasmodiu
35	70.6	4.2	1442	7	MTSDATP9A	MTSDATP9A Plasmodiu
36	70.4	4.2	15861	37	AE001413	AE001413 Plasmodiu
37	70.2	4.2	2857	36	DDIPSTCAT	DDIPSTCAT Plasmodiu
38	69.8	4.2	222401	34	PFMAL3P3	PFMAL3P3 Plasmodiu
39	69.8	4.2	312766	34	PFMAL3P3	PFMAL3P3 Plasmodiu
40	69.2	4.2	1910	7	SCMAJ4148	SCMAJ4148 Saccharom
41	69.2	4.2	220475	35	AC006278	AC006278 Plasmodiu
42	69	4.1	163141	34	HS1110P6	HS1110P6 Homo sapi
43	69	4.1	4601	36	DMU11584	DMU11584 Drosophila
44	69	4.1	19517	37	DMU37541	DMU37541 Drosophila
45	68.8	4.1	110000	34	PFMAL4P1_1	PFMAL4P1_1 Continuation (2 of

#### ALIGNMENTS

RESULT 1	BB062900	1655 bp	DNA	BCT	15-JAN-1997
LOCUS	Borrelia burgdorferi	flagellar filament outsheath protein (flaA)			
DEFINITION	gene, complete cds, and chemotaxis histidine kinase (cheA) gene, partial cds.				
ACCESSION	U62900				
NID	51575445				
VERSION	U62900.1	GI:1575445			
KEYWORDS					
SOURCE					
ORGANISM	Borrelia burgdorferi				
	Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.				
REFERENCE	1 (bases 1 to 1655)				
AUTHORS	Ge, Y. and Charon, N.W.				
TITLE	An unexpected flaa homolog is present and expressed in Borrelia burgdorferi				
JOURNAL	J. Bacteriol. 179 (2), 552-556 (1997)				
MEDLINE	97144545				
REFERENCE	2 (bases 1 to 1655)				
AUTHORS	Ge, Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (03-JUL-1996) Yigong Ge, West Virginia University,				

Microbiology, HSCN, Morgantown, WV 26506, USA  
Location/Qualifiers  
1. 1655  
/organism="Borrelia burgdorferi"  
/strain="212"  
/db\_xref="taxon:139"  
/clone="pW1 and pW2"  
1. 395  
/function="unknown"  
/note="Oria"  
/codon\_start=3  
/transl\_table=1  
/protein\_id="AAC44769.1"  
/db\_xref="PID:g1575446"  
/db\_xref="GI:1575446"  
/translation="DNLSKVFYFHSKGFVANLRYLRDQNLKNDLILVDFLLG  
SNEGSEFLSDRLFLSKNGVYVYNSREFYDFNNGDYNEFVDVYNLFAMSL  
IKTRFNPGRKIKIILVSGCILKEGS"  
473. 1498  
/gene="flaA"  
473. 1498  
/gene="flaA"  
/function="putative flagellar filament outsheath protein"  
/codon\_start=1  
/transl\_table=1  
/product="FlaA protein"  
/protein\_id="AAC44770.1"  
/db\_xref="PID:g1575447"  
/db\_xref="GI:1575447"  
/translation="MKRAKSLFELLSTVLFNQTDLAEGSKRAEPGVLDPAFL  
ARPSSTRDLTNYDVYVSGASIVPEPMVYDLGINNMSVLTPEBARQAYKNVY  
VAPVASESKRYAGDTILGKRVLPSTYSSASATMPKIPITSGSGNQFJGKLI  
DNITMKEIVSYSLGVEIDLEVLFEIDMGMEYAYSMGLKFGMDLIMSPNTIP  
NISRIIHKDVPNPPLASMKRFKAFVSKSVKSNFIFVYDLKLVYDKLSVSD  
SDISESVFVYFERSGESLRIKLAHETFKRYLKRKISIAEGSFNFYEKIESEK  
EESPKN"  
1566. 1655  
/gene="cheA"  
1566. 1655  
/gene="cheA"  
/function="chemotaxis histidine kinase"  
/codon\_start=1  
/transl\_table=1  
/product="CheA protein"  
/protein\_id="AAC44771.1"  
/db\_xref="PID:g1575448"  
/db\_xref="GI:1575448"  
/translation="MEILDENEELGVFFPEAQNVLILEENI"  
BASE COUNT 560 a 182 c 308 g 605 t  
ORIGIN  
Query Match 98.4%: Score 1637; DB 2; Length 1655;  
Best Local Similarity 99.5%: Pred. No. 4.2e-196;  
Matches 1655; Conservative 0; Mismatches 0; Indels 8; Gaps 1;  
QY 1 atgabaatccttttcaaaaaaggtttttatcattcctagcaaggattgtgcta 60  
DB 1 ATGATAATCTTTTCAAAAAAGGTTTATTTTCATTTCAGCAAGGATTGTGCTA 60  
QY 61 atttaagatatttaagaagaatgaaacaattgaaagaataattagatctttaagaag 120  
DB 61 ATTTAAGATATTATTAGAAATGAACAAATTTGAAAGATTAATTAGATCTTTAGTAAAG 120  
QY 121 atttctcttagaagcaatgagggcttctcttgggtttttattatgaagtatcaaat 180  
DB 121 ATTTCCTTTAGGAAGCAATGAGGGCTTTCTTTGGGTTTATTAAAGTCATTCAGAT 180  
QY 181 ttattatctcttttaagaagaatggagttattatgtaaatctttcaagaagaattttat 240  
DB 181 TTTTATATTCTTTTAAAGATGAGTATTATGTAATCTTTCAAGAGAAATTTTATG 240  
QY 241 attctttaataaagtgatataatgaatcttttgatgatttaaggatcaatcttttgcta 300

|||||  
DB 241 ATTCTTTTAATTAATGATTAATGAATCTTTTGATGTTAAGGCAACTTTTGCTA 300  
QY 301 tgtcttaataaacaatcgcttactatcctcgtgtaagaataaataattatttc 360  
|||||  
DB 301 TGTCTTTAATTAATTAATGATTAATGAATCTTTTGATGTTAAGGCAACTTTTGCTA 360  
QY 361 ttgttgaagggtgtaatccttaagaagaatgtaataatttaataatttaataa 420  
DB 361 TTGTTGAAGGCTGATCTTAAGAGCAAGTGTAAATTAATTAAATTTACTAATAA 420  
QY 421 aattaaaaacgaaattttataaagatttatataaagagttggttactgtaaa 480  
DB 421 AATTAAAAACGAAATTTTATAAAGATTATATATAAGAGTGGTTTACATGAAG 480  
QY 481 gaaagcaaaagtatttatttttttattatcactcagttcttttgcataagaact 540  
DB 481 GAAAGCTAAAGATTTTATTATTTTATTATATCACCAGTCTTTTCTCAAGAGACTGA 540  
QY 541 tggattagcagaagggtcttaaaaggcagacccgagaatagctttagatttgcga 600  
DB 541 TGGATTAGCAGAGGCTCTAAAGGCGACGCTGAGAAATTAGTCTTATGATTTGCGCA 600  
QY 601 gcttgcaagagatcccaagttcaactagactgattcttaacaattatgtattatgata 660  
DB 601 GCTTGCAAGAGATCCCAAGTCAACTAGACTGATCTTACAAATTAATGATTATGATA 660  
QY 661 ttccggagctcttgtaattgtaaccgagaatatggttgatgacttggtgataa 720  
DB 661 TTCCGGGCGCTTGTGATTTGTTAAGCGGAGATATGTTGATGATTTGGGATAAATA 720  
QY 721 ttgagcgctttacttactcctcctcgtcaaggttgcaggttcaagtttaaaatcagttgt 780  
DB 721 TTGAGCGCTTTACTTCTCTCTCTGCAAGGTTGACAGCTTACGTTAAATTCAGTTGT 780  
QY 781 tggcccgctggttgaagggtgataaaggtaagcgagtgatactatttgggggt 840  
DB 781 TGCCCGCGCTGTTGATTAAGGTGAGTCAAAAAGGTAGCGAGGTGATATTTTGGGGGT 840  
QY 841 aagagcttcttccaaagcttccatcatcctgctataattatgacacatttaaat 900  
DB 841 AAGAGTTTGTTCGAAGCTATTCATCATCTGATATATATATGACACATTAAAT 900  
QY 901 tccctttattcagggaagggtgcaatcaattttaggaagaagttctatgataacat 960  
DB 901 TCCCTTTATTTCAGGCGAAGGTGCAATCAATTTTAGGCAAGGCTTATGATTAACAT 960  
QY 961 taacaacatgaagaatattagatctgtttatagtttaggttagatgaatatttga 1020  
DB 961 TAAACCATGAAGAAATTAAGGATCTGTTATAGTTTAGGATGAGATATCTTGA 1020  
QY 1021 gttttattgaagataatgagcatgataatgcttattcattatggttactttaa 1080  
DB 1021 GCTTTATTGAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
QY 1081 taaagggtgggtcgtatatttattgtaaaatcctaactatattcctaataatc 1140  
DB 1081 TAAAGGTTGGGCTGATTTAATTTGTCAAATCCAACTATATATTCCTAATATCATCAG 1140  
QY 1141 aattattaaagaagatgttccaaattatcctctgctcctaagaataaagtgat 1200  
DB 1141 AATTATTAAAGACATGTTCCAAATTTCCCTGCTTCAAGTAAATGAGATTTAAGGC 1200  
QY 1201 tttaagagtttcaagtcacacagttcaaaagtttaaaatttcatcttttatt 1260  
DB 1201 TTTTAGGTTTCAAGTCACACAGTTCAAAAGTTTAAATTTTCTTTTATGTTAAGA 1260  
QY 1261 tttaagagtttcttattgataagcctaaggtttcaatagattctgataatgacagttg 1320  
DB 1261 TTTTAGGTTTCTTATGATTAAGGTAAGTGTTCATATGATTCGATTTGACAGTGATC 1320  
QY 1321 tgtatctaaagttatgagcctagcggaactgaatcccttcgtaaattaaggcacaga 1380  
|||||

Db 1321 TGTATTTAAAGTTATATAGACAGTACGGAACGTAATCCCTTCGTAAATTAAGGACACAGA 1380  
QY 1381 aactttaaagaagctttaaagcttagagaaaaattctcagctgaagctcttcca 1440  
Db 1381 AACTTTAAAGAGCTTTAAAGCTTAGAGAAAATTTCTATCCCTGAGGCTTTTCCA 1440  
QY 1441 aaactttagaagaagatctgagagtgaaaaaaccctgaagaatcatctccgaaaaataggt 1500  
Db 1441 AAACCTTGTGAGAAAAGATGTGAGAGTGAAGAAACCTGAGAAATCATCTCCGAAAATTAAGT 1500  
QY 1501 ttaactaatatgtgaagcttaacaaagcttgctctterroysnncatatttaataa 1560  
Db 1501 TTAATTAATGTAAAGCTTAAAGCTTAAAGGTTTCCTT-----ACATATTAATAATAA 1552  
QY 1561 taggaatagatatacgaaatattagatttggaaaaatgaagagctttagaggttttt 1620  
Db 1553 TAGGAATAGATATGTGAAGAAATTTAGATTGGAAAATGAAGAGCTTTTAGAGAGTTT 1612  
QY 1621 tgaagaagctcaaaaactctgtagatatccttgaagaagaatatt 1663  
Db 1613 TGAAGAGCTCAAAAATCTGTAGATATCCTTGAAGAGAAATATT 1655

RESULT 2  
AE001168 13271 bp DNA BCT 15-DEC-1997  
LOCUS Borrelia burgdorferi (section 54 of 70) of the complete genome.  
DEFINITION AE001168 AE000783  
ACCESSION g2688598  
VERSION AE001168.1 GI:2688598  
KEYWORDS  
SOURCE Lyme disease spirochete.  
ORGANISM Borrelia burgdorferi  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group  
1 (bases 1 to 13271)  
Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A., Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K., Gilm,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D., Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S., Hanson,M., Van-Vugt,R., Palmer,N., Adams,M.D., Gocayne,J.D., Weidman,J., Uterback,T., Watthey,L., McDonald,L., Attich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K., Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi  
Nature 390 (6660), 580-586 (1997)  
JOURNAL 98065943  
MEDLINE 2 (bases 1 to 13271)  
REFERENCE Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A., Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K., Gilm,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D., Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J., Salzberg,S., Hanson,M., Van-Vugt,R., Palmer,N., Adams,M.D., Gocayne,J.D., Weidman,J., Uterback,T., Watthey,L., McDonald,L., Attich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K., Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
TITLE Direct Submission  
SUBMITTED (12-DEC-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
JOURNAL Location/Qualifiers  
FEATURES  
source 1. 13271  
/organism="Borrelia burgdorferi"  
/db\_xref="taxon:139"  
gene 66. 1100  
/gene="BB0668"  
CDS 66. 1100  
/gene="BB0668"  
/note="similar to GP:1575447 percent identity: 98.83; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="flagellar filament outer layer protein (flaA)"

gene 1168. 3762  
/gene="BB0669"  
CDS 1168. 3762  
/gene="BB0669"  
/note="similar to PID:1113815 SP:044737 GB:AE000783 percent identity: 99.31; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis histidine kinase (cheA-2)"  
/protein\_id="AAC67024.1"  
/db\_xref="PID:g2688607"  
/db\_xref="GI:2688607"  
/translation="MEIDLLENELLGVFEENQNLVDILEENIMSLIEDDPNNSDTID EIPRAHTLKGSASIDMELSDFTFHVIEDVEPAIDRGVNNINDVLISLDVIK EMALRIDGKRYLNDISDLSKLOFLVDDOFFIKRFGNSIKNNFCESDLEIR EGIIGOKVLRISVFNNSNSSEVNSGKIFNLKNGSVLHTIKRYQIIEDEKRL RNDYIYIYSDIBSVKSLDSNLIESYLVDERVNEKELKLADEIKVDIDSNFPLN DNFDEDEISDLLENQKLVKRLVDKRPMPATISGLQMLQKLSKGTFSIP DSSELDADKDFEVIYLLISNTSESIARKNIPDVVSFEIKNNVLESKSVRLKD DEAFENKRNKNSPISVNLIRIDSKDIYILNISEAVISKSNQINSEMTITLFY NENFYDYDOESFORNLIDLVKFDAGKLIEDEISHINSLASFMEKALNISLAR NSPEFLLQNKMSGRSLRITIDHESVLTBLPISNIFRSTRVRRVRLKNTIKV NIKMESEELPELGSVIDDVLDPMLHVRNSMNGLETVEERVNRGSKAGTILKRN EGVNISTEIEDGIDGIDPVIRKRLKEGKIKDAISDEELNIFAFGEFSAVOT DLSRGVGLDVAKRSIEKNGTILVSESGIGTIFIKLPLTLVIYIQLGKVSGETY VIPLNVLHTRHTEIDKLENTYHEVYNLRODVISVLRLDKLFTNTRDLSLEKFLI VVTSNMKKAIAVDSILGDEDFVVRPIKDKFSSAGIVGATTLGNGKVVLLIDVFKLF DLQDKRKE"  
3769. 5169  
/gene="BB0670"  
CDS 3769. 5169  
/gene="BB0670"  
/note="similar to GP:1881576 percent identity: 98.93; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="purine-binding chemotaxis protein (cheW-3)"  
/protein\_id="AAC67023.1"  
/db\_xref="PID:g2688606"  
/db\_xref="GI:2688606"  
/translation="MOKIEYRGPKILDDKNSKLTNDFKVSVEELGSDHYLVDM OVEKIRKSNFTYVPAKRYVALDMLREIPIIDILRINFENPKKIDIELMVKR EDLILVIDIKNNVSIDSLSLODHPYELIINIKGVADYEKELIILNVKRI FNGEKEKILKPCONVEKSDVSDCDNDILENSKNSKNSKNDGNEKSTLNN TAFENLEKRNKLNKVSFNASLVNDVLELVGKFMVNDVDYPCDSPLNEFYKSGG NIMGACCTEEKREKIEVSRSPNNNNSITFNVEICSGCKETMALANALSEYVVP KIRALINDLSKVEYTRLVFSESEIGISIEISNRSQSGVITKLSISNVLFEIS DALFSDLPNLGVETIRLNDVLCFLDSKDOIILINILASKALKGALLVLGDNELKNDV FIKRKSTYFNLRKY"  
5181. 5666  
/gene="BB0671"  
CDS 5181. 5666  
/gene="BB0671"  
/note="similar to GP:1408273 percent identity: 98.76; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis operon protein (cheX)"  
/protein\_id="AAC67022.1"  
/db\_xref="PID:g2688605"  
/db\_xref="GI:2688605"  
/translation="NRIDYIEPLDAASVLRDMLVENVEMCKPGLKSNOKIGVS





Db 901 ATTCTGATATTGACAGTACTGTGTATTAAAGTTATGAGCTAGCGGAACTGAATCCC 960  
 QY 1359 ttctgaattaaagcagcagaaactttaaaagagttctaaagcttagaagaaaaattt 1418  
 Db 961 TTGTAATTAAGGCGACAGAACTTTAAAAGAGTTTAAAGCTTAGAGAAAAAATTT 1020  
 QY 1419 ctatgcctgaaggctcttccaaaacttgcagaaaagattgcagtgaaaacctgaag 1478  
 Db 1021 CTATGCTGTAAGGCTCTTTCCAAAACCTTTAGAGAAAGATTGAGAGTGAAGAAACCTGAG 1080  
 QY 1479 aatcattctccgaaaaaatagggttaataatagttaaaactactctaaaggtttgctt 1538  
 Db 1081 AATCATCTCCGAAAAAATAGGTTTAATATATATGTAACCTAAAGGTTTGCCTT 1140  
 QY 1539 erroysnacataataaataatagaaatagatataatgaaatattagattgaaatg 1598  
 Db 1141 -----ACATATTAATAAATATAGAAATAGTATATGAAATATTTGATTGGAAATG 1192  
 QY 1599 aagagcttttagagagtttttttgaagaagctcaaaactctgtagatatccttgaaaga 1658  
 Db 1193 AAGAGCTTTAGAGAGTTTTTTGAAGAACTCAAAATCTGTGATATCTTGAAGAGA 1252  
 QY 1659 atatt 1663  
 Db 1253 ATATT 1257  
 RESULT 3  
 AEO01167  
 LOCUS AEO01167 10016 bp DNA BCT 15-DEC-1997  
 DEFINITION Borrelia burgdorferi (section 53 of 70) of the complete genome.  
 ACCESSION AEO01167 AEO00783  
 NID 92688585  
 VERSION AEO01167.1 GI:2688585  
 KEYWORDS  
 SOURCE Lyme disease spirochete.  
 ORGANISM Borrelia burgdorferi  
 Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
 burgdorferi group.  
 1 (bases 1 to 10016)  
 Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,  
 Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
 Gwinn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,  
 Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,  
 Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D.,  
 Gocayne,J.D., Weidman,C., Garland,S., Fujii,C., Cotton,M.D., McDonald,L.,  
 Artlich,P., Bowman,C., Hatcher,B., Smith,H.O. and Venter,J.C.  
 Genomic sequence of a Lyme disease spirochete, Borrelia  
 burgdorferi  
 Nature 390 (6660), 580-586 (1997)  
 JOURNAL 98065943  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 TITLE  
 Submitted (12-DEC-1997) The Institute for Genomic Research, 9712  
 Direct Submission  
 Medical Center Dr. Rockville, MD 20850, USA  
 FEATURES  
 source  
 1..10016  
 /organism="Borrelia burgdorferi"  
 /db\_xref="taxon:139"  
 84..1217  
 /gene="BB0656"  
 84..1217  
 /gene="BB0656"  
 /note="similar to SP:P54304 PID:1303805 GB:AL009126  
 /note="similar to SP:P54304 PID:1303805 GB:AL009126

percent identity: 25.62; identified by sequence  
 similarity; putative"  
 /codon\_start-1  
 /transl\_table-11  
 /product="oxygen-independent coproporphyrinogen III  
 oxidase, putative"  
 /protein\_id="AAC67009.1"  
 /db\_xref="PID:92688591"  
 /db\_xref="GI:26885891"  
 /translation="MRVDLPVLESLVYNLSFCCDSIFRTILELKHILILGHP  
 IKTIXIKHVDCLSRQDNKIFETSLSKYINLELEETLEITIGYDDEFKILDE  
 FCTIRINLNVOSFSLERFVIGFPISTKLKILINIKKEFFDLNIDTVMEQSLQK  
 SHLKDLKELSLYMPBHCFSDFICEEGVLRDSDSDSEKLFACLECSGYI  
 NYEITNFALKGHESRHNKLMELKPHLGLYAVSLDFCNDNNVRAILIRKTSFVK  
 ANNHVLTPELLEDEFYVHFIFOGGTIOGVLRALRLREVENEQFOFIYVCSLTS  
 KKFVDDNMLIKGRRERKNFYLVKIINHFNDFKYLRLP"  
 complement(1204..1890)  
 /gene="BB0657"  
 complement(1204..1890)  
 /note="similar to GB:L77117 PID:1500500 percent identity:  
 39.53; identified by sequence similarity; putative"  
 /codon\_start-1  
 /transl\_table-11  
 /product="ribose 5-phosphate isomerase (rpi)"  
 /protein\_id="AAC67008.1"  
 /db\_xref="PID:92688590"  
 /db\_xref="GI:26885890"  
 /translation="MENQKILYAKRAIDHYIKSNMNLGIGTITTYIAIKYSEIKS  
 GSKLNFETYSDDRYLLSKQIDPESEFSKLNKNDLDAIDGAEILLEKSLKGM  
 GGAHLMERKVIANSSTLLIADETRIYKLGKMPIDIEVAQNAVGFIETLEENLE  
 ATLRIKKEKKEPTIDNNYIIDVAKMHWENPEGTREYKFLPGILEGIFNHNKTRIV  
 YXDDKOIKXA"  
 2024..2785  
 /gene="BB0658"  
 2024..2785  
 /note="similar to SP:P31217 GB:U00096 PID:1651340  
 PID:1786970 percent identity: 59.76; identified by  
 sequence similarity; putative"  
 /codon\_start-1  
 /transl\_table-11  
 /product="phosphoglycerate mutase (gpmA)"  
 /protein\_id="AAC67007.1"  
 /db\_xref="PID:92688589"  
 /db\_xref="GI:2688589"  
 /translation="MGDPKILVLRHGESWKNENLFTGTYDKLSKGDIAVEA  
 GILLQEGISFDIAFSSLSRANDPLNLIIRLEGOSYSIVKTRWLNERNHYALGLN  
 KSETAKYGEDRVLIWRSYDVPMPSLSDSDRHPIDKDPYKHIPRELSPYELKDT  
 VARVIPYMTDEIAKEVLEGGKVIYVAHGSRLALVYFDNLSEEDVLKINPTGIPV  
 YELDNLNPIKHYVYGDESKIKKAMESVASQGLK"  
 complement(2853..4418)  
 /gene="BB0659"  
 complement(2853..4418)  
 /note="similar to GB:L77117 SP:057959 PID:1591243 percent  
 identity: 33.14; identified by sequence similarity;  
 putative"  
 /codon\_start-1  
 /transl\_table-11  
 /product="lysyl-tRNA synthetase"  
 /protein\_id="AAC67006.1"  
 /db\_xref="PID:92688588"  
 /db\_xref="GI:2688588"  
 /translation="MKTAMADFYAEKIKKEGPKNLYTVASGTPSGTVHGNFV  
 ISVDLVARLRDGSKVRPIYSGMDNDVDFKPKNPEDELTLTLRQILITVPDTRS  
 HRTSYARANEIEFEKYLPGVGINPEIDDSQYTSNAYASQIKFALDHKELESLANE  
 YRTSLKEWNPYISVECTKNDRTTVNNVDHYSVEYCEGNGESLDIRTTMAIKL  
 PRRIDMPMKRYEKEDPEAGDHHSSGSPSTKNIVYIFGSGSPVTVQYDPSIKG  
 RGKRTSSSGDVISLKDVLVYTPPEYTRPLPAATRPNTEFSISFDIDYKITEIDKF  
 ERTIYGEVDEKEERKRAFRITELSOPIWPSKRIPIYQVGFRLSVISQIFENNINKL  
 NTLKNVQDQDKRLINKINCAINWIRDFEDFKSLRSKFPNMEILENSKKAANEL  
 LDFLKNFEVATEODIQNEIKYISRENNIEPALFFKOYIKILIDKEKGPRLAGFIKI

gene  
CDS  
GIDREFKITSKYV  
/complement(4505..5377)  
/gene="BB0660"  
/complement(4505..5377)  
/note="similar to GB:M4658 SP:P06616 PID:416295 PID:42767 PID:987641 percent identity: 37.85; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="GTP-binding protein (era)"  
/protein\_id="AAC67005.1"  
/db\_xref="GI:2688587"  
/db\_xref="GI:2688597"  
/translation="MKSQFALIGRPSNGSKTLLNSIGKHSIISPIPORTNNG  
IFPDROGFIIDPFGPHLSKKKNIMMKYHSISGEVELIYIDIDOKPGEERK  
MELIKNSKIFELYLNKIDAKNTKKEITFOLEKGEEDSNITIKISAEKINTEELK  
NKIYENFSEGFLYPOEYTDQETNFISETIRKALENLEELPISLYVIDITLENK  
KGSLEIRANIFVANSQKGIIVNGREKISIGERRAKRTAKIRETRCNELQVLEK  
NMNKEDLIKRLIN"  
5467..5838  
/gene="BB0661"  
5467..5838  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0661"  
/protein\_id="AAC67015.1"  
/db\_xref="GI:2688597"  
/db\_xref="GI:2688597"  
/translation="MRNFRALKVNFEEVCECGEKGVLNRSVCNCAKLGIDICPC  
RHVGVPAGFGCPCNCHSPROELKEKPFKKEARVAVSDAAAKTVRLHEFQIND  
ILYLFSSFLPILFLYIVG"  
5848..6258  
/gene="BB0662"  
5848..6258  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0662"  
/protein\_id="AAC67014.1"  
/db\_xref="GI:2688596"  
/db\_xref="GI:2688596"  
/translation="MGITTFYLSIFASFYVGSMSDYKKNVTKSTIFYDVEVERP  
YARKQTLQFIKTLKTAVERNENKMSYTFVDKLIQYALFIEVKRFGKATLV  
TPLNVLMDLGDISIIVLNKILIRILKSYISMYNK"  
6278..6727  
/gene="BB0663"  
6278..6727  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0663"  
/protein\_id="AAC67013.1"  
/db\_xref="GI:2688595"  
/db\_xref="GI:2688595"  
/translation="MKRIKRFLLFLVLSMSPFLSPADHFDYDEKIMINGVKEFKIASNE  
LDRKGYMGAQKVEYNGMLPVEKKDONLSFMMENMTPLLEIAVYIDSGIIEKYDLE  
PYSRANYSILKRYKALELPRGSKFOIKIGDVYHCFDYNLSLVE"  
6736..7419  
/gene="BB0664"  
6736..7419  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11

gene  
Query Match 26.2%; Score 435.8; DB 2; Length 10016;  
Best Local Similarity 97.6%; Pred. No. 9.1e-47;  
Matches 456; Conservative 0; Mismatches 2; Indels 9; Gaps 1;  
/product="B. burgdorferi predicted coding region BB0664"  
/protein\_id="AAC67012.1"  
/db\_xref="GI:2688594"  
/db\_xref="GI:2688594"  
/translation="KLHNTWLIRININNTLFPGMIIIFALISCHKNIOYRKIK  
FLDKNIEKIDSENDPFAEKDINNEKEEVLIRSLNYSKSKIEIFGIYVFDIN  
TPRIKESIDSLMSDSYNNRFGSMELIHNAERINSIVYIVKAEPAANTFLDAIDE  
IASITISFKRIITNNENIDNNENNNTNESNQPLKOEKTSRESNNELKQDIE  
EELQEIKAQ"  
7462..8421  
Qy 1 atgataatcttttcaaaaaaggtttttatcttcttctagaaaggttgctcta 60  
|||||  
Db 9550 ATGATTAATCTTTTCAAAAAAGGTTTTTATTTTCATTCAGCAAGGATTTGTTCTTA 9609  
Qy 61 attaagatatcttaagagatgaacaaattgaaagataattagatcttcttaaaag 120  
|||||  
Db 9610 ATTTAAGATATTTAAGAGATGAACAAATTTGAAGAATATTGATCTTTAGTTAAAG 9669  
Qy 121 atttcttttagaagaagaatgaggggtttcttcttggttttaataagtgatcaaat 180  
|||||  
Db 9670 ATTTCTTTTGAAGCAATGAAGGTTTCTTTGGGTTTTTATTAAGTATCAAGAT 9729  
Qy 181 ttatatctctttttaagaagatgagtttatcttgaatcttcaagaagatttatg 240  
|||||  
Db 9730 TTTTATATCTTTTAAAGATGAGATTTATATGTAATCTTTCAAGAGAAATTTATG 9789  
Qy 241 attctttaataatgctgatla-----taatgaatctttgatgcttaagctcaatc 291  
|||||  
Db 9790 ATTTTAAATTAAGGATATTAATGAACATATGAATCTTTTGATGATTAAGTCAATC 9849  
Qy 292 ttttgcctatgctttaataaaacaatgcgctttaactcctgtaagataaaaaa 351  
|||||  
Db 9850 TTTTTCATGCTTTTATATAAAACAATGCGCTTAACCTGCTGAAGATAAAAAAGA 9909  
Qy 352 ttatatctcttggtaaggggtgtatcttaagaaggaagtgtataaatttact 411  
|||||  
Db 9910 TTGTATTCCTTTGGAAGGGGTATCTTAAGGCAAGTGTAATTAATTAATTTACT 9969  
Qy 412 aataaaataaataaaacgaataatttaataagaattatata 458  
|||||  
Db 9970 AATATAAATATTAATAAACGAATTTTATATAAGATTATATATA 10016  
RESULT 4  
BCHEAGEN  
LOCUS BCHEAGEN 332 bp DNA BCT 20-JAN-1998  
DEFINITION B. burgdorferi chea gene.  
ACCESSION X91907  
NID X91907.1  
VERSION X91907.1 GI:1019753  
KEYWORDS chea gene; chea protein; chemotactic response protein; histidine kinase.  
SOURCE Lyme disease spirochete.  
ORGANISM Borrelia burgdorferi  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia burgdorferi group.  
REFERENCE 1 (bases 1 to 332)  
AUTHORS Trueta, G.A., Old, J.G., Saint Girons, I. and Johnson, R.C.  
TITLE A chea cheu operon in Borrelia burgdorferi, the agent of Lyme disease  
JOURNAL Res. Microbiol. 148 (3), 191-200 (1997)  
MEDLINE 98438936  
REFERENCE 2 (bases 1 to 332)  
AUTHORS Old, J.G.  
TITLE Direct Submission  
SUBMITTED (03-OCT-1995) I.G. Old, Institut Pasteur, Bacteriology & Mycology, 28 rue du Dr Roux, 75724 Paris cedex 15, FRANCE  
FEATURES location/Qualifiers

```
source 1. 332
        /organism="Borrelia burgdorferi"
        /strain="212"
        /db_xref="taxon:139"
        /clone_lib="Cos1"
gene 1. 202
        /gene="orf"
        <1. 202
CDS /gene="orf"
        /codon_start=2
        /transl_table=11
        /protein_id="CA63001.1"
        /db_xref="PID:g1019754"
        /db_xref="GI:1019754"
        /db_xref="SPTREMBL:Q44876"
        /translation="DSDIDSESVFKVYETSGTESLRKKAHETFKRYLKREKISINAE
        GSFOHFKEIKESKEPSSPKN"
repeat_region 205. 256
        /note="hairpin"
RBS 258. 261
gene 270. 332
        /gene="cheA"
        270. >332
CDS /gene="cheA"
        /function="chemotactic response"
        /codon_start=1
        /transl_table=11
        /product="histidine kinase"
        /protein_id="CA63002.1"
        /db_xref="PID:g1019755"
        /db_xref="GI:1019755"
        /db_xref="SWISS-PROT:Q44737"
        /translation="MEILDENELLCVFEPAQN"
BASE COUNT 130 a 37 c 61 g 104 t
ORIGIN

Query Match 18.98; Score 314; DB 1; Length 332;
Best Local Similarity 97.68; Pred. No. 4.1e-31;
Matches 332; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1297 agattcgattatgaacagtgctgtatcttaagaattatgaactagcgaactgaatc 1336
    |||||||
DB 1 AGATTGATATTTGACACGAGTCTGATTTAAAGTTATGAGACTACCGAAGCTGATC 60

QY 1357 ccttcgtaaatataagagcagcaactttaaaagctttaaagcttagaagaaaaaat 1416
    |||||||
DB 61 CTTTCGTAATTAAGGACACGAACTTTAAAGAGTTTAAAGCTTAGAGAAAAAT 120

QY 1417 ttctatcgctgaaggtctcttccaaaacttctgtagaagaattgagagtgaaaaacctga 1476
    |||||||
DB 121 TTTATCGCTGAAGGCTCTTCCAAACTTTGTGAAAGATTGAGAGTGAAGAAACCTGA 180

QY 1477 agaatacattccgaaaaaattggtttaaataatgtaaaagcttaaaaggtttgct 1536
    |||||||
DB 181 AGAATCATCTCCGAAAAATTAGTTAAATTAATGTAAAGCTAACCTAAAAAGTTTCT 240

QY 1537 tterroysmncatataaataataggaataatagatataatagattatagttgaaaa 1596
    |||||||
DB 241 TT-----ACATATTAATAAATAAGGAAATAGATATGAAAAATTAGATTGGAAAA 292

QY 1597 tgaagaagctttagagatttttttgaagaagctcaaat 1636
    |||||||
DB 293 TGAAGAGCTTTAGGAGTTTTTTTTGAAGAGCTCAAAAT 332

RESULT 5
LOCUS BBU28962 2700 bp DNA BCT 16-JAN-1998
DEFINITION Borrelia burgdorferi histidine kinase (cheA) gene, complete cds.
ACCESSION U28962
MID 9113814
VERSION U28962.1 GI:1113814
```

```
KEYWORDS
SOURCE Lyme disease spirochete.
ORGANISM Borrelia burgdorferi
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
burgdorferi group.
REFERENCE 1 (bases 1 to 2700)
AUTHORS Trueta,G.A., Old,I.G., Saint Girons,I. and Johnson,R.C.
TITLE A cheA cheW operon in Borrelia burgdorferi, the agent of Lyme
disease
JOURNAL Res. Microbiol. 148 (3), 191-200 (1997)
MEDLINE 98438936
REFERENCE 2 (bases 1 to 2700)
AUTHORS Trueta,G.A. and Johnson,R.C.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1995) Gabriel A. Trueta, Microbiology, University
of Minnesota, 420 Delaware Street S.E., Minneapolis, MN 55455, USA
COMMENT On Dec 12, 1995 this sequence version replaced gi:984552.
FEATURES
    source
        1. 2700
            /organism="Borrelia burgdorferi"
            /strain="CT-1"
            /db_xref="taxon:139"
            95. 2689
            /gene="cheA"
            95. 2689
            /gene="cheA"
            /codon_start=1
            /transl_table=11
            /function="Involved in chemotaxis"
            /product="histidine kinase"
            /protein_id="AB96835.1"
            /db_xref="PID:g1113815"
            /db_xref="GI:1113815"
            /translation="MEILDENEELIIVFEPAQNVLIDLEINMSLEDDPNNSDTID
            EIRRAATLKGSASIDMMELSPFTIIVEDVPATIDGKYNINNDVLIDLSLDPIK
            EMALATIDGKVYIINDISDKSLKQFLVIDDQTFIRFDNSIKNNFCJSESDLEIR
            EGIGIQKVLRIIVFNSSNSNHEVNGKIFENIKNLGSLVLTIRKRYQIJDIEKFLK
            RVDYVLIYSIDIEGVKSLDSNLIESYLVDDEFVKEKLADEIKVDLIDKIDFYN
            DNDPFEDEISDLLEVENOKLFKVRIDFKDNPMTISGLQLOAKSLGKFKGISP
            DSEELADKFEEDVIYIILISNTSEESISKINLPDVSHPEIKVNLVLSIKSVRLKD
            DEAPFKENRIKNSPIYVNLIDSKIDYIINLVSEAVISKSTNOINSEVITLTFY
            NENFYDYDSFGKRNFLIDIKIVFKDAGLLEELISHNSLSFMEKALKOISEIR
            NSPFLQNKMTSGRLSRITDLHESVKTBMPLSNIESRTRVARDLPKLNKIV
            NLMEGEETEELDSVIDLDVDPIMHCVRNSMDGLTEVERVYRGSKAGTILRAKN
            EGNVISIEIEDIGIDIPYIRKLIIEKGIKEDAIYSOELINLIPAPGFTAVOYT
            DLGKRVGLDYYVKKSIKELNGTILVSEIGLGTFIKKLPVTIITIGLVKSGSEY
            VILNNVLEHRIHEDIKLENTHEVYINLRDVISVLRIDKLFNITRDSDLIEKFLI
            VVNTSMKLAIVVDSILGEDFVVKPIKDKFSSAGIVGATTLGNGRVYLIIDVFKLF
            DLODKYKE"
BASE COUNT 988 a 278 c 507 g 927 t
ORIGIN

Query Match 9.98; Score 164.4; DB 2; Length 2700;
Best Local Similarity 95.38; Pred. No. 8.8e-13;
Matches 183; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 1472 cctgaagaatcattccgaaaaattagtttaatatatgtaaaagctaacctaaagct 1531
    |||||||
DB 1 CCTGAAGAATCATCTCCGAAAAATTAGTTAAATTAATGTAAAGCTCAAAAGCT 60

QY 1532 ttgctterroysmncatataaataataggaataatagatataatagattatagttg 1591
    |||||||
DB 61 TTGCTTT-----ACATATTAATAAATAAGGAAATAGATATGAAAAATTAGATTG 112

QY 1592 gaaaaatgaagagctttagagatttttttgaagaagctcaaaactctgtagatatcct 1651
    |||||||
DB 113 GAAATGAAGAGATTTAGAGGTTTTTTTGAAGAAGCTCAAAATCTGTAGATATCCTT 172

QY 1652 gaagaagaattt 1663
    |||||||
DB 173 GAAGAGATATTT 184
```

	RESULT	6			
	LOCUS	THFLA1	1260 bp	DNA	BCT
	DEFINITION	T.hydroysenteriae flaI gene.			
	ACCESSION	X63006	S38343		
	NID	G433523			
	VERSION	G433523	GI:433523		
	KEYWORDS	flaI gene; FlaI protein; flagellar sheath; flagellin; periplasmic protein.			
	SOURCE	Brachyspira hyodysenteriae.			
	ORGANISM	Brachyspira hyodysenteriae			
	REFERENCE	Bacteria; Spirochaetales; Spirochaetaceae; Brachyspira.			
	AUTHORS	1 (bases 1 to 1260)			
	TITLE	Kusters,J.G.			
	JOURNAL	Direct Submission			
	FEATURES	Submitted (04-NOV-1991) J.G. Kusters, Inst. of Infectious Diseases and Immun., P.O.Box 80.165, 3508 TD Utrecht, THE NETHERLANDS			
	REFERENCE	2 (bases 1 to 1260)			
	AUTHORS	Koopman,M.B., de laeuw,O.S., van der zeljst,B.M. and Kusters,J.G.			
	TITLE	Cloning and DNA sequence analysis of a Serpulina (Treponema) hydroysenteriae gene encoding a periplasmic flagellar sheath protein			
	JOURNAL	Infect. Immun. 60 (7), 2920-2925 (1992)			
	MEDLINE	92307926			
	FEATURES				
	source	location/Qualifiers			
		1..1260			
		/organism="Brachyspira hyodysenteriae"			
		/strain="C5"			
		/db_xref="taxon:159"			
	-35_signal	27..32			
	-10_signal	50..55			
	RBS	86..91			
	gene	103..1065			
		/gene="flaI"			
		103..159			
	sig_peptide	/gene="flaI"			
		103..1065			
	CDS	/gene="flaI"			
		/codon_start=1			
		/transl_table=11			
		/product="flagellin FlaI"			
		/protein_id="CAA44735.1"			
		/db_xref="PID:g433524"			
		/db_xref="GI:433524"			
		/db_xref="SMSS-PROT:P32520"			
		/translation="MKRLFVLTSIFIAASAAYGLTNSLIDFALTGNAANDNGEPT NEYPAENLYNDWVYWNLSALTERRRSYTNDSEKNRAMEAGYLGRVHFF PLAMNSYALKPYELEMTGGAAGITETBEKGVIHNVEIKTSSVYGNVLIYSF VNLONEFGELKSYPMGTYVENGWQVRNKEGYLPNWDRLVREPLYPRMIPSVKLDSLIGYRTKDTKGDFITYVKDVTLEYVVVDVFEDIEDDEATWOLLKTENDRKQAIESSE ARIREOALERDLEQRIRIGDGTAAADQGAANTGAADTGAAGBOAQ"			
	mat_peptide	160..1062			
		/gene="flaI"			
		/product="flagellin FlaI"			
	repeat_region	1004..1030			
		/rpt_type="INVERTED"			
	repeat_region	1157..1175			
		/rpt_type="INVERTED"			
	BASE COUNT	437 a 158 c 245 g 420 t			
	ORIGIN				
	Query Match	6.5%; Score 108.6; DB 1; Length 1260;			
	Best Local Similarity	49.4%; Pred. No. le-05;			
	Matches 369; Conservative	0; Mismatches 369; Indels 9; Gaps 3;			
Oy	581	ttagctttagatatttcgacggacttgcaagagataccaagtccaactagactgatctaca	640		
Db	136	TTTATCGGTGCATCGTCGTACGGTTTACAACAACCACTTGATTGATTTCCTTACA	195		
Oy	641	aactatgttatcatatcatatccgcgccttcggtatattgttaaccgcgaagatatgctt	700		
Db	196	GGTATGCTGTAATAACTTACAACTGAGAGAGGTGATACAAAAGAATGATTCCAAGTTCA	255		

[illegible]

sig\_peptide /note="putative"  
 CDS 103..159  
 103..1065  
 /codon\_start=1  
 /transl\_table=11  
 /product="endoflagellar sheath protein"  
 /protein\_id="CA01715.1"  
 /db\_xref="PID:g904393"  
 /db\_xref="GI:904393"  
 /transl\_except="P32520"  
 /translation="MKLFVLTSLIFSLASAVGLTSLIDPALTGNDNLQAGEBDT  
 NEVPVPAENLYNDNNVYVWLNESARLTENRNSYTVTVNDGNGAEGVLRVHF  
 PLAAVSVLVKPVYELVEMVGGADGTRKGVINVGKISISSWYPMRPSVYLD  
 VMLQNEFGEKSYPMGVYNGRQVWRENRREYLPVVRDLREPLVPMRPSVYLD  
 SLGFYRKDKGSDFTYVADVLEEDVYVDEEDIDDEATLQIKTENDRQALIES  
 ARIRQAEIRDLERIRIGDGTADQGAANTGAADGAAEQNQ"

BASE COUNT 390 a 152 c 233 g 365 t  
 ORIGIN  
 Query Match 6.5%; Score 108.6; DB 5; Length 1140;  
 Best Local Similarity 49.4%; Pred. No. 1e-05;  
 Matches 369; Conservative 0; Mismatches 369; Indels 9; Gaps 3;  
 QY 581 ctgagcttagatttcgcgagcttcgaagatccaaagttcaactagactgacatcaca 640  
 DB 136 TTATCGCTCATCTCTGCTTACGCTTAAACAACCACTTGTATGATTTGCTTAA 195  
 QY 641 aattatgtatcatatgataatttcggcgcttcgttattgtaagccggaagatagct 700  
 DB 196 GGTATGCTGATTAACCTTACAGCTGGAGAAAGTATCAATGAGTAGTTCAGTTG 255  
 QY 701 gtagacttgagaaataaattcgagcgctttactactcccttcgaagcttcagct 760  
 DB 256 GAAATCTTATATGATGATGCTGATGATGATGATGATGATGATGATGATGATG 315  
 QY 761 taagcttaaaattcagttgctgcgcgcgtgctgtaagagtgagcaaaaggatcaca 820  
 DB 316 AATCGCGAATATCTTATGCTTACTACGTAGACAGTAAAGGTAACAACGCTTGG 375  
 QY 821 ggtgatactatttggggggaaggtttgttccaaagcattctc---aactctgct 877  
 DB 376 GCAAGTAAAGTCTTGGTGTAGAGTACATTTCCCATTTAGCAGCTTGGACAGTAA 435  
 QY 878 atgatacgcaaccattaaattccctttattcaggagggaagtggaatcaatt--- 935  
 DB 436 TTAGTAAACACAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 495  
 QY 935 ctaggcaaaagcttattgtaacattaaacacgaagaagaattgaagatctgttat 994  
 DB 496 GAAGGTAAAGGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 555  
 QY 995 agtttaaggtatgagatctgaggtttatttgaagatagaattgcaatgcaat 1054  
 DB 556 GGACGTAACTATTTAATTTAGTATTTGCTAACTTACAAATGAAATTTGGAATTA 615  
 QY 1055 gcttactcttggactcttaaggttcaaggttcgaagctgatttaattgtcaaatc 1114  
 DB 616 TCTTATCAATGGTACTGTTTACTTCAACGCTTGGAGACAGTAAAGTAAAGTAA 675  
 QY 1115 aactatactcctaataatacaaccagaattattaaagacagatgtccaaattact 1174  
 DB 676 GAATACTTACTTATGTTCCGACAGAGTATTTAGTAAAGAACTCT---TTATC 732  
 QY 1175 gcttcaagtaaaatgagatttaaggtttcagagttccaaagtcacacagttcaa 1234  
 DB 733 ATGATCCCTCTCTGTTAAATTTAGTATTTAGTATTTAGTATTTAGTATTTAG 792  
 QY 1235 aaaaattcttcttataagatttaagaggttcttattatgataagtaaggtttca 1294  
 DB 793 GGAATTTTCACTTACCTTAAAGATGTAACTTGAAGTATGACGTAGTAGTGTG 852  
 QY 1295 atagattctgataattgacagtgagct 1321

DB 853 TTGGAAGACATATCGACGATGAAGCT 879  
 RESULT 8  
 AC006986/c 173956 bp DNA PRI 05-MAY-1999  
 LOCUS Homo sapiens clone NH0155J05, complete sequence.  
 DEFINITION AC006986  
 ACCESSION 4753246  
 NID AC006986.2 GI:4753246  
 VERSION HTG.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 173956)  
 AUTHORS Waterston,R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 173956)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-MAR-1999) Genome Sequencing Center, Washington  
 REFERENCE 3 (bases 1 to 173956)  
 AUTHORS University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE 4 (bases 1 to 173956)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-MAY-1999) Genome Sequencing Center, Washington  
 REFERENCE 5 (bases 1 to 173956)  
 AUTHORS University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 COMMENT On May 5, 1999 this sequence version replaced gi:4337255.  
 FEATUERS  
 source  
 1..173956  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="NH0155J05"  
 BASE COUNT 54090 a 33708 c 33792 g 52366 t  
 ORIGIN  
 Query Match 5.4%; Score 89.4; DB 42; Length 173956;  
 Best Local Similarity 49.9%; Pred. No. 0.0005;  
 Matches 225; Conservative 0; Mismatches 226; Indels 0; Gaps 0;  
 QY 60 aatttaagataattgaagatgaacaatttgaagataattagatctttagtaaaa 119  
 DB 2123 AAATTAAT 2064  
 QY 120 gatcttcttgaagaaagcaatgaggtcttcttgggtttattatgaagttcaaga 179  
 DB 2063 AT 2004  
 QY 180 ttcttaattctttttaaagaatggagtttattatgtaaatcttcaagagaatttat 239  
 DB 2003 ATATTAAT 1944  
 QY 240 gatcttcttgaagaaagcaatgaggtcttcttgggtttattatgaagttcaaga 299  
 DB 1943 AT 1884  
 QY 300 atgtctttaaataaacaagcgcttaactaccggtaagataaaagattattatt 359  
 DB 1883 AT 1824  
 QY 360 ctgttgaaggtgatctttaaaggaagagtttaatttaatttaatttaatttaatt 419  
 DB 1823 TATGAATAAT 1764  
 QY 420 taattaaagaagaatttataaagattatataataagaggttggtttatagaaaa 479  
 DB 1763 TATATGAATAAT 1704



RESULT 11  
LOCUS PFMAL3P7 253307 bp DNA INV 29-APR-1999  
DEFINITION Plasmodium falciparum MAL3P7, complete sequence.  
ACCESSION AL034559  
VERSION 94723992  
KEYWORDS GI:4725992  
SOURCE HTG.  
ORGANISM malaria parasite P. falciparum.  
REFERENCE 1 (bases 1 to 253307)  
AUTHORS Oliver, K., Bowman, S., Harris, D., Lawson, D., Quail, M. and Barrell, B.  
TITLE Direct Submission  
JOURNAL Submitted (15-APR-1999) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK  
COMMENT On Apr 29, 1999 this sequence version replaced gi:4493963.  
For more information about this sequence or the Malaria Project, see <http://www.sanger.ac.uk/Projects/P.falciparum>. IMPORTANT: This sequence is not the entire insert of clone MAL3P7. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

FEATURES  
source Location/Qualifiers  
1. .253307  
/organism="Plasmodium falciparum"  
/strain="3D7"  
/db\_xref="taxon:5833"  
/chromosome="3"  
/clone="MAL3P7"  
458. .4060  
/gene="MAL3P7.1"  
458. .4060  
/note="predicted using hexexon; MAL3P7.1 (PFC0860w), kinesin-related protein, len: 1201 aa; Similarity to kinesin-related protein, D.melanogaster kinesin-like protein 67A (TR:P91945) BLAST Score: 666, sum P(2) = 1.9e-67; 288 identity in 707 aa overlap."  
/protein\_id="CAB39023.1"  
/db\_xref="PID:e1407858"  
/db\_xref="PID:g4493964"  
/db\_xref="GI:4493964"  
/translation="MKTKNLNDPFLNNTYKKEENPIIGSDERKRTKITYSTNH  
HTNNILNDHTLKLICNITENKALHSKSNLQKNITLMPYQKGGHIRETKN  
VYENINGEPTSEVENTNGEHFINGQYDALKNLSLNNDHNNINNNISNKKRLEVL  
NNNNNNNDVIEENMEYPTISKNIYDSIYIPOINTIKNIINSEINNNNNINDNNH  
NYTNIHNNNNNSNPTSSVTSKNNNNNLINTLANSKVAVRIRKIGSESENYS  
IFKNYVILKEKEKCYLLSOKKOSTYFDVFPYNATOEVEFFOTAPFLIPHYK  
GINCTYFATGATSGKTYTMDLKNONGIVQLSLELFTIINERKRNKIVLSLEY  
YNTITDLGKEKNTLEVOEAEVKSNCIEVNNTEQAMLINEGYKRNKMSPT  
RAKVSRSRHAIQIIVYNEIIDDNNMTISYKALCFVDLAGSERASANSKGERKE  
GSYINSLALACINSLASNRNISKYVRKDSKTLHLKNSLEGNCLVMTANIP  
SRSPQSNNTLKCIAFRARIRKICATVOTRDNKESDIEKIKNNENLQREYDILRY  
TKKEPFIINVTQIKKOISCTIKLEINSDMSSMELKODITHTDQVKMSDEIR  
KAVDSLKDIQEOKFLNLDTFLERKNLNTVINSKVDNNNSLSEMLFEHNNK  
VKNELVNEKVVKNVNGVAVDENVDKNNVAVDENVAVDENVAVDENVAVDENV  
VAVDENVAVDENVAVDENVAVDENVAVDENVAVDENVAVDENVAVDENVAVDEN  
ADESSHNKNNKDIEDNDKDTIKDINHNNNSDNDDEYOSANSVPSDIYKREK  
KTIPIMETKKRTNGCTDPIHKTPTDITNGITLNKEDVSNKSNYNTKATERNKY  
EKKEGVNPHNNITLDMONSLIYIINNENSHSPRKNVAKMLKGLNLTANLIL  
DDDTLNMMNSKISDHNKNSNLIINNENKINDSKCKNINNINNINNINNINN  
NNNNNNSSSSSSKVDGINLNNSTNELHTFSYSLNLDEIKIEILNKDMKKN  
DHLTIDITSKHARDLKENKRKLENFQENIKHHEKDEVSLYVKKKIKKKK"  
5457. .6743  
/gene="MAL3P7.2"  
/note="predicted using hexexon; MAL3P7.2 (PFC0865w),  
/gene="MAL3P7.2"  
CDS  
gene  
CDS

RNA-binding protein, len: 310 aa; Similarity to C.elegans  
RNA-binding protein, C.elegans RNA-binding protein  
(TR:Q18318) BLAST Score: 378, sum P(1) = 3.8e-35; 37%  
identity in 297 aa overlap."  
/codon\_start=1  
/protein\_id="CAB39067.1"  
/db\_xref="PID:e1407902"  
/db\_xref="PID:g4494008"  
/db\_xref="GI:4494008"  
/translation="MGLFDKIRINIEKLNELKNGNDSNHHDOYRDSSTIYIGLND  
NLTESDIYVFSOYGEPIDVNLVNDENEGSKGCFISYAORSILVAVDNPNGYKL  
LERPLVYDILITVRLPKTKLADKDKERKPTABEGOGITVYNNVEISLTFYFKIK  
NKSNEKKRKLIDDELMLNLEKSLIKDIIISPIGDESRNHEGKEEDEDDED  
DDEDDDDSDVIDIKYKRNRKRSLSITTKYDKRKRKRSRDRKRRRRENSRHRKX  
EKDKSKRRHRHSDKYSRSRSTSSSTDR"  
7801. .8583  
/gene="MAL3P7.3"  
/note="predicted using hexexon; MAL3P7.3 (PFC0870w),  
Elongation factor 1-beta, len: 181 aa; Similarity to  
elongation factor 1-beta, A.salina elongation factor  
1-beta (SW:EF1B.ARTSA) BLAST Score: 159, sum P(2) =  
4.9e-16; 488 identity in 62 aa overlap."  
/codon\_start=1  
/protein\_id="CAB39068.1"  
/db\_xref="PID:e1407903"  
/db\_xref="PID:g4494009"  
/db\_xref="GI:4494009"  
/translation="MANTYDELVPLSYIILONEGNTSKIDQANTKPKKEVINKS  
SLIIDIKPYGENTDDEVLKLVKNITMEGLTGAKHKRPPAFGLQVYNNKRYI  
TSIYIYIYVYMLFEVSCIVVDLVNTDELITETIENLGLDNEQLQKKQMDDEYD  
EDDEIGLVOSAEIISFNKL"  
11431. .20442  
/gene="MAL3P7.4"  
/note="predicted using hexexon; MAL3P7.4 (PFC0875w),  
Hypothetical protein, len: 3004 aa"  
/codon\_start=1  
/protein\_id="CAB39024.1"  
/db\_xref="PID:e1407859"  
/db\_xref="PID:g4493965"  
/db\_xref="GI:4493965"  
/translation="MKDFSYAENNDINTFKLYNSPEDCLNKLKVAIEKSPKVN  
SLENDIINTIENHEDKSLNLYNHTIINTLKSEPIINDLYKNTYENFENIKIKI  
NPHILKVANNTEFTKNEFINDKLNKENNTIYTFENDRKLEFYNVKNNLVETKISC  
GLSVENVYKRNKQKVPKSYLGLSPQEGKGGKGLVYNAPLYNKEKQKGNPFL  
YYPVYLEDKRIENKERNKERNKERNKERNKERNKERNKERNKERNKERNKERN  
NGIOSYHNNNSINSNDOKKNNNNNNNYNNNDGLINNVYKIVGAYALNSNEK  
YEPNDINILKONFENFTVVDLSFTYVNNENYSEFVLEOFNSPFIYANADILK  
KNMLTPRDEINELKEKEITMEKKTDEKEQVNNETINENVGIDEGGYKTELK  
LNDFTYKPKMSMKINADTEKNIERVAVELCYCLFVNICFDINKERKINTEFL  
COLKIKYIYYSWLFEPYILFEFNYILFLVYFVYVACLNLYFLFLIYFFLNS  
LLEFTILCOMFNNSSINYIATLFLFESSFLRILVSSVNLITFLVLIIPSSCLS  
LFTFLILNNIKIDKOLFKEPNSLHLHILIGSVSVNLICLNTIYIYKRRKM  
MNTKVRGSKSTIDREKRSKSOOKTEYENIYDMSGEPYQLODNSDPMLESGDE  
DTCYIEDTLKQVNSQCLNTEKTRNDRESNRPDIPIMLPHANNINNINNINN  
INNINNINNINNINNQHNNOCYLLITQVNVKPYGKHALKOVSLTRKSNRLVYLGE  
NSGKCTLNIITIKMSKSGEINFEKNSYKKNKKDIYNDVKNRNGLCGRICSGY  
IKFEVNRNFKIRKNDNTNNDNSNNNNNNNDHNSNGSINSEFNNNNVNNKI  
KNDLEISYQVNVYILEYETIKIPELLYNNKVDYLLNKKRKLKINNNDVLOYL  
NNOIKRLMEKVKKISIEFLCYLRDIIYILDEPFIADIKTKTLKFPFKDIKKNNIT  
FICHDITVANNPANDIAIKSGOILIFKSKSFOLEIKYKTLNIOFGNLNEEDIT  
YLSQERTSALKKRGVITKPRVANNYVGVVGRISSESSINDKKONNDNNDD  
NNKNSVDHIDHFLDYININEFTNNDIIRKNNIKIIFKTRNENKCFIEFNEHDI  
YCTYKINELSESKLCLCYNKEKNIYOLKTDIYTYIYITLNEKKKLKNIODK  
DKYLCIEIOLPEFLFQNFENLNNLILKNNHOPISYNEAYRDIOLVNGENK  
IINSTNDNDPYDINYNHKKYKNNNNNSPYNMPDDELDIENISSKKN  
KTEKEFSINTYRNKKNDENDGCTYDNDNMISNTSNMKKRNKQNIPTCI  
NEFTYIKPTLLKRLKDDGSSSFYWKIVPLFLSFLLLIKCVSLFGKQVNDLID  
YTSISSNHLKOSTINFIYKNNMNNNELGYNNVNLNDNDIKNNLNSYDKILKLE  
TKODNAEESHINNNDNDNIKDILKENNEYNNKLHENSFNYSNTINSLOKVC





Query Match	4.6%	Score 77	DB 35	Length 256172
BASE COUNT	100700 a	26900 c	25826 g	101937 t
ORIGIN	809 others			
FEATURES	source 1..256172 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /chromosome="12"			
COMMENT	On Apr. 2, 1999 this sequence version replaced gi:4337170. * NOTE: This is a 'working draft' sequence. It currently * consists of 5 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. 1 126535: contig of 126535 bp in length * 126536 126735: gap of unknown length * 126736 158845: contig of 32110 bp in length * 158846 159045: gap of unknown length * 159046 211538: contig of 52493 bp in length * 211539 217738: gap of unknown length * 217739 248106: contig of 36368 bp in length * 248107 248306: gap of unknown length * 248307 256172: contig of 7866 bp in length. location/Qualifiers			
REFERENCE	AUTHORS Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W. TITLE Direct Submission JOURNAL Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA REFERENCE 1 (bases 1 to 256172) AUTHORS Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W. TITLE Plasmodium falciparum 3D7 chromosome 12 JOURNAL Unpublished			
LOCUS	AC005139 256172 bp DNA HTG 01-APR-1999			
DEFINITION	Plasmodium falciparum chromosome 12, WORKING DRAFT SEQUENCE, 5 unordered pieces.			
ACCESSION	AC005139			
NID	64558581			
VERSION	AC005139.3 GI:4558581			
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	malaria parasite P. falciparum.			
ORGANISM	Plasmodium falciparum			
REFERENCE	1 (bases 1 to 256172) Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
AUTHORS	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
TITLE	Direct Submission			
REFERENCE	Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA			
AUTHORS	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
TITLE	Plasmodium falciparum 3D7 chromosome 12			
JOURNAL	Unpublished			
LOCUS	2 (bases 1 to 256172)			
DEFINITION	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
ACCESSION	AC005139			
NID	64558581			
VERSION	AC005139.3 GI:4558581			
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	malaria parasite P. falciparum.			
ORGANISM	Plasmodium falciparum			
REFERENCE	1 (bases 1 to 256172) Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
AUTHORS	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
TITLE	Direct Submission			
REFERENCE	Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA			
AUTHORS	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
TITLE	Plasmodium falciparum 3D7 chromosome 12			
JOURNAL	Unpublished			
LOCUS	2 (bases 1 to 256172)			
DEFINITION	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
ACCESSION	AC005139			
NID	64558581			
VERSION	AC005139.3 GI:4558581			
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	malaria parasite P. falciparum.			
ORGANISM	Plasmodium falciparum			
REFERENCE	1 (bases 1 to 256172) Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
AUTHORS	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
TITLE	Direct Submission			
REFERENCE	Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA			
AUTHORS	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
TITLE	Plasmodium falciparum 3D7 chromosome 12			
JOURNAL	Unpublished			
LOCUS	2 (bases 1 to 256172)			
DEFINITION	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
ACCESSION	AC005139			
NID	64558581			
VERSION	AC005139.3 GI:4558581			
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	malaria parasite P. falciparum.			
ORGANISM	Plasmodium falciparum			
REFERENCE	1 (bases 1 to 256172) Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
AUTHORS	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
TITLE	Direct Submission			
REFERENCE	Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA			
AUTHORS	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
TITLE	Plasmodium falciparum 3D7 chromosome 12			
JOURNAL	Unpublished			
LOCUS	2 (bases 1 to 256172)			
DEFINITION	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
ACCESSION	AC005139			
NID	64558581			
VERSION	AC005139.3 GI:4558581			
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	malaria parasite P. falciparum.			
ORGANISM	Plasmodium falciparum			
REFERENCE	1 (bases 1 to 256172) Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
AUTHORS	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
TITLE	Direct Submission			
REFERENCE	Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA			
AUTHORS	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
TITLE	Plasmodium falciparum 3D7 chromosome 12			
JOURNAL	Unpublished			
LOCUS	2 (bases 1 to 256172)			
DEFINITION	Hyman,R.W., Fung,E.L., Qiu,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.			
ACCESSION	AC005139			
NID	64558581			
VERSION				







GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 16:24:08 ; Search time 148.09 seconds  
(without alignments)  
2809.571 Million cell updates/sec

Title: US-09-004-395-1  
Doct# 1663

Sequence: 1 atgataatctttttcaaaa.....atatccttgagaagaataatt 1663

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36:★

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Query	Match	Length	DB	ID	Description
No.	Score					
C 1	1588.4	95.5	110000	1	X20248_01	Continuation (2 of
C 2	1588.4	95.5	110000	1	X20248_02	Continuation (3 of
C 3	107	6.4	1140	1	Q38563	43 kD endoflagellu
C 4	78	4.7	19124	1	T72882	Plasmodium var-7 g
C 5	67	4.0	6124	1	Q03568	Sequence encoding
C 6	64.8	3.9	3975	1	M81155	Malaria-specific g
C 7	64.8	3.9	3975	1	Q22959	SEK1 gene. Recomb
C 8	58.4	3.5	4590	1	M60472	Sequence encoding
C 9	57.8	3.5	1612	1	M60392	Sequence encoding
C 10	57.6	3.5	2418	1	Q27866	P.falciparum GPI1
C 11	57.6	3.5	2104	1	Q25273	Sequence encoding
C 12	57.2	3.4	9789	1	T41852	CDNA encoding Plas
C 13	56.8	3.4	6124	1	Q03568	Sequence encoding
C 14	56.8	3.4	605	1	T31530	Human 3' apolipop
C 15	56	3.4	8920	1	Q62974	Carbamoyl-phosphat
C 16	55.6	3.3	605	1	T31530	Human 3' apolipop
C 17	55.4	3.3	4673	1	Q27189	P. yoelii SSP2 anti
C 18	55.4	3.3	5849	1	W33155	Plasmodium berghei
C 19	55	3.3	5760	1	N50530	Sequence encoding
C 20	55	3.3	19124	1	T72882	Plasmodium var-7 g
C 21	54	3.2	4818	1	T79909	Actin gene. Promote
C 22	54	3.2	4818	1	T62117	Candida boidinii A
C 23	53.6	3.2	1864	1	N71405	Sequence of ANS 1
C 24	53.6	3.2	110000	1	X20248_06	Continuation (7 of
C 25	53.6	3.2	111309	1	X20250	Borrelia burgdorferi
C 26	53.2	3.2	1297	1	Q21645	3' coding sequence
C 27	53.2	3.2	1770	1	Q37684	P.falciparum antiad
C 28	53.2	3.2	1770	1	Q37124	Borrelia burgdorferi
C 29	53.2	3.2	4228	1	X20277	Sequence encoding
C 30	52.6	3.2	3101	1	Q02047	Sequence encoding
C 31	52.6	3.2	4590	1	M60472	Sequence encoding
C 32	52.4	3.2	2763	1	N95079	Gmbs26-A heat sho
C 33	52.2	3.1	9789	1	T41852	CDNA encoding Plas
C 34	52	3.1	1982	1	N90225	Malaria-specific E
C 35	51.8	3.1	4838	1	T70891	Plasmiid pTR-cep-1a
C 36	51.6	3.1	2503	1	Q53460	pMPX30 xylanase c
C 37	51.6	3.1	2401	1	X20301	Borrelia burgdorferi
C 38	51.6	3.1	110000	1	X20248_08	Continuation (9 of
C 39	51.4	3.1	3600	1	T77330	Solanum tuberosum
C 40	51.4	3.1	53585	1	X20251	Borrelia burgdorferi
C 41	51.2	3.1	5760	1	N50530	Sequence encoding
C 42	51.2	3.1	1511	1	Q28302	AMEPV tk DNA. New
C 43	51.2	3.1	1511	1	Q66798	AMEPV thymidine-k

## ALIGNMENTS

C	44	51.2	3.1	1511	1	V14508	AMEPV entomopoxvir
C	45	51	3.1	2418	1	Q27886	P. falciparum GBP13
ALIGNMENTS							
RESULT 1							
X20248_01/c							
Continuation (2 of 10) of X20248 from base 100001 (Borrelia burgdorferi polynucleotide							
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248							
WP	Fragment Name			Begin	End		
WP	X20248_00			1	110000		
WP	X20248_01			100001	210000		
WP	X20248_02			200001	310000		
WP	X20248_03			300001	410000		
WP	X20248_04			400001	510000		
WP	X20248_05			500001	610000		
WP	X20248_06			600001	710000		
WP	X20248_07			700001	810000		
WP	X20248_08			800001	910000		
WP	X20248_09			900001	910715		

Query Match	95.5%;	Score 1588.4;	DB 1;	Length 110000;
Best Local Similarity	98.2%;	Pred. No. 1.2e-257;		
Matches 1644;	Conservative 0;	Mismatches 11;	Indels 19;	Gaps 3

OY	1	atgataacccctttccaaaaaggcttcttatcttcctcagaagaaggcttggtgcta	60
Db	105623	ATGTAATACCTTTTCCAAAAAGCTTTTATTTTCACTTCTACAGACGATTTGTTCCTA	105564
OY	61	attcaagatcttaagatgagacaaacttgaaagalaatcttagatcctttagtaaaag	120
Db	105563	ATTTAAGATATTATTAGAGATGACAAAATTGGAAGATATTATAGACTTTTGTAAAG	105504
OY	121	atttcctcttgagagaacatgaggagcttctcttggtgtttcttaagtgatccaagat	180
Db	105503	ATTTCCTTTTGGAGACATCAGAGGCTTTCTTTGGGCTTTTATTAGTGATTCAGAT	105444
OY	181	tttatactctcttttaagaagatgagttctattatgtaaaccttcaagaagactttaig	240
Db	105443	TTTTATATCTTTTAAAGAAATGAGATTATATGTAAATCTTTCAGAGAAATTTATG	105384
OY	241	attccttttaataaagggtgatta-----taatgaactcttttgatgctaaagctaac	291
Db	105383	ATTCTTTTAAATAAAGGGATTTATATAGTAATCTAATGTGAATCTTTTGAGTTAAGCTCATC	105324
OY	292	ttttgtcctatgctcttataaacaatgcgcttaactcctccggtgaagataaaaaaga	351
Db	105323	TTTTTGCATGTCCTTTATATAAAACATGCGCTTAACCTACCTCGGTAAAGTAAAAAGA	105264
OY	352	ttatctatctctgttgaaaggctgatactttaaaggagcaagctgataaataaatttact	411
Db	105263	TGTGTATCTCTGTGAAGGGGTGTCTTTAAAGGCAAGTGTATAATTAAATTTTACT	105204
OY	412	aataaaaataattbaaaaacgaanaatttttaagaagt--tatatrtaaaggagtgtgtc	469
Db	105203	AATATAAATAATTAAAAACGAAATAATTATAAAGAATTTATATATATTAAGAGAGTTGGTT	105144
OY	470	tacatgaaaagaagatctaaagaattttactttcttatatccactgttctcttgcct	529
Db	105143	TACATGAATAAGGAAAGCTAAAGATATTTATTTTATTTATATACACTGTCTTTTGGCT	105084
OY	530	caagagactgtagattagcagaaggtctctaaaaggcagaagccttgaagaatagctcta	589
Db	105083	CAAGAGACTGTAGGATTAAGCAGAGGGCTCTTAAAGGGCAGAGCCTGGAGAAATTAGTTTA	105024
OY	590	gatttgcgcgagcttgcaagagatccaagtccaactgagacttgatcttcaaaattatgtc	649
Db	105023	GATTTTCCGAGGCTTGCAGAGATCCCAAGTCCACACTGAGACTGTATCTTACAAATTATGTT	104964
OY	650	gattatgataatctcgggcctctctgtatgttgaagccggaagatalggttgttagacctt	709

```

|||||
Db 104963 GATTAATGATATTCGGCCCTTCGTGATGTTAAAGCCGGAAGAAATGGTTAGATCTT 104904
OY 710 gggataaataatgagcgcttactactcctctgcaaggctgaagcttaagtttaa 769
|||||
Db 104903 GGGATTAATAATGAGCGCTTTACTACTCCTTCGCAAGGTTGACAGCTTAACGTTAAA 104844
OY 770 aatcaagtgtgtgcgcgcgtgtgttaagagtgagtcacaaagagtagcaagtgatact 829
|||||
Db 104843 AATTCAGTTGTGCGCCCGCTGTGTTAGTAGTGAATCAAAAAGTAGCCAGTAGTACT 104784
OY 830 atttgggggttaagagtttgtttccaaagctattccaaatcgtcgtctatgtatgcca 889
|||||
Db 104783 ATTTTAGGGGTAGAGTTTGTGTTCCAAAGCTATTCGAATCAATCGCTATGATTAAGCCA 104724
OY 890 ccatttaaatcccttttattccagggaagtggaacatcaattttaagcaaaagttct 949
|||||
Db 104723 CCATTTAAATTCCTTTTATTTCAGGGGAAAGTGGCAATCAATTTTAAAGGCAAGGCTTT 104664
OY 950 attgataacattaaacacalgaagaanaatlaagatctgtttatagtttaaggtatgag 1009
Db 104663 ATTTGATTAACATTTAAACCATGAAGAAATTAAGTATCTGTTTATAGTTAGGTATGAG 104604
OY 1010 atagatcttgaggttttatttaagaataatgagatgagcaatgagtaattctatgggt 1069
Db 104603 ATAGATCTTGAGGTTTATTTTGAAGATGAATGGCATGGAATATGCTTATCTATGGGT 104544
OY 1070 accttaaaagtttaaaaggtggtgcatttaatttgttcaaatcccaataatcttctaatt 1129
|||||
Db 104543 ACTTTTAACTTTTAAAGGGGGGCGCTGATTTAATTTGGTCAAACTTACATTAATTTCTTAAT 104484
OY 1130 atataccacagaataattaaagacagatgtccaaatctcctctgtcttcaagtaaatg 1189
Db 104483 ATATCATCCAGATATTTATTAAGACATGATGTTCCAAATATATCTCTGCTCAAGTAAAG 104424
OY 1190 agatttaaggtctttagagtttcaaaagtcacacagtttcaaaagtttaaaattctatctt 1249
Db 104423 AATTTTAAAGGCTTTTGAAGTTTCAAAAGTCCACAGTCAAAAGACGCAAAATTTCAATCTTT 104364
OY 1250 tatgttaaaagatttaagagttcttataagtaagctaaaggtttcaaatatctatatt 1309
Db 104363 TATGTTAAAGATTTAAGAGTTTCTTATGATTAAGTTGAGTGGCTTCAATGATTTCTGATATT 104304
OY 1310 gacagtgagctgtatcttaaaagttataagagctagcagagctgaatcccttgcataat 1369
Db 104303 GACAGTGAAGTCTGTATTAACTTATGAGCTAGAGGAACTGAATCCCTTGCTGAATTA 104244
OY 1370 aaggaacccgaaccttttaaaagagttttaaagcttaaggaanaaattctctcgtgaa 1429
Db 104243 AAGGCAACAGAACTTTTAAAGAGTTTAAAGCTTAGAGAAATTTCTATGCTCGTA 104184
OY 1430 ggccttttccaaacctgtgtagaaagattgagagtgaaacacgaagatcatctcgt 1489
Db 104183 GGCTCTTTCCAAACACTTCTTCTATGATTAAGTTGAGTGAAGGAAACCTGAACATCACTCCG 104124
OY 1490 aaaaattaggtttaaatlaataatgtaaaagctacctaanaaggttgcctterroysnaa 1549
Db 104123 AAAAATTTGGTTTAAATTAATATGTAAGCTACCTAAAGGTTGGCTT-----ACA 104072
OY 1550 tatttaataataggaataatagatatggaataatatactagatttggaaaaatgaagagctttta 1609
Db 104071 TATTTAAATAATAGAAATAGTATATGAAATATTAAGATTGGAAATTAAGAGCTTTTAA 104012
OY 1610 ggaagtttttttgaagaagctcaaaactctgtagatatccttgaagagagatact 1663
Db 104011 GGAGTTTTTTTTGAAGAAGCTCAAAATCTTTGATATCTCTTGAAAGAGAAATATT 103958

```

RESULT  
 X20248\_02/c  
 Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide s  
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
 WP Fragment Name Begin End

```

WP X20248_00 1 110000
WP X20248_01 100001 210000
WP X20248_02 200001 310000
WP X20248_03 300001 410000
WP X20248_04 400001 510000
WP X20248_05 500001 610000
WP X20248_06 600001 710000
WP X20248_07 700001 810000
WP X20248_08 800001 910000
WP X20248_09 900001 910715

```

Query Match 95.5%; Score 1588.4; DB 1; Length 110000;  
 Best Local Similarity 98.2%; Pred. No. 1.2e-257;  
 Matches 1644; Conservative 0; Mismatches 11; Indels 19; Gaps 3;

```

OY 1 atgataaccttttccaaagagtttttttatttcttcaatccagcaagaggtttgtcta 60
Db 5623 ATGATTAATCTTTTCAAAAAAGTTTATTTTCATTCAGCAAGGATTTTGCTA 5564
OY 61 attaaagatttaagagatgaacaaatttgaagaataattagatctttagtaaaag 120
Db 5563 ATTTAAGTATTTTAAGAGATGACAAATTTGAAGATTAATTTAGATTTAAG 5504
OY 121 atttcttttaggaagcaatgaggggttctcttcttggtttttaaagtgatccaagat 180
Db 5503 ATTTCTTTTGAAGAACATGAAGGTTTCTTTGGGTTTATTAAGTATTCAGAT 5444
OY 181 ttatatctctttttaaagaatgaggtttatatgtaaatccttcaagagaatttatg 240
Db 5443 TTTTATATCTTTTAAAGAGATGAGTTTATGTAATCTTTCAAGAAATTTATG 5384
OY 241 attcctttaaatggttacta-----taatgaatcctttgagtttaagttcaatc 291
Db 5383 ATTCCTTTTAATTAAGGTATTTATTAATGATATGATTAATCTTTTGATGTTAAGTCAATC 5324
OY 292 ttttgtcagtcctttaaataaacaaatgcgtttaaatacctcgtgtaagaataaaaga 351
Db 5323 TTTTGTGATGCTTTTATTAATAAACAAATGCGCTTAACATCTGTAAGATTAATAAAGA 5264
OY 352 ttatatctctgttgaagaggtgtatcttaaaggcgcaagttgataaatttaattact 411
Db 5263 TTTGTTATCTTTTGAAGAGGTTGATCTTTAAAGGCAAAAGTTGATTAATTAATTTAACT 5204
OY 412 aataaataatlaaaaaacgaataatttaaaagatt--tatataagaggttggt 469
Db 5203 AATTAATAATTTAAAAACGAATTTTATTAATAAGATTTTATATATTAAGAGATTTGCTT 5144
OY 470 taatgaaaggaagctaaagaatttatttttttatatccactgcttcttltgct 529
Db 5143 TACATGAATAAGAAAGCTAAAGATATTTATTTTATTAATCCAGTCTTTTGGCT 5084
OY 530 caagagactgtatgattatagcagaaggttcttaaaggcgagagccttgagaatttgctta 589
Db 5083 CAAGAGACTGATGATTAAGCAGAGGCTTTAAAGGCGAGACCTGGAGATTTGTTTTTA 5024
OY 590 gatttgcgagcttgcagaagatccaaagttcaactgaacttgatcttaacaattatggt 649
Db 5023 GATTTTGGCGAGCTTGCAAGAGATCCAAATGCACTAGACTTGATCTTACAAATTAATGTT 4964
OY 650 gattatgataatcgggagcctctcgtgatttgttaagccggaagatatggtttagatcct 709
Db 4963 GATTATGTAATATGCGGGGCTTCTGATGTTTAAAGCCGAAGATATGTTGTAAGATCTT 4904
OY 710 gggataaataatgagaggttttactactcctctcgaaggttgcaaggttcaagtttaa 769
Db 4903 GGGATTAATAATTTGAGGCTTTACTTACTCTCTGCAAGGTTGACAGGCTTTAGCTTAAA 4844
OY 770 aatcaagtgtgtgcgcgcgtgtgtttaaagatgaagtaaaagagtagcagaggtatct 829
Db 4843 AATTCAGTTGTGCGCCCGCTGTGTTTAAAGTATGATCAAAAGAGTACGACAGGATGATCT 4784
OY 830 atttgggggttaagagtttgtttccaaagctattcctaactatcgtcatatgataatgcca 889

```

```

Db 4783 ATTTAGGGGTAGAGATTGTTTCAGATATTCACATCTCTATGATATGACA 4724
QY 890 ccaattaaattcctttttatcaggagggaagtggcaatcaatttttagcaaggctt 949
Db 4723 CCATTTAAATTCCTTTTATTATTCAGGGGAAAGTGGCAATCAATTTTAGCAAGGCTT 4664
QY 950 atgataacataaaacacagaaagaatbaagtaactgtttatagtttaggtatgag 1009
Db 4663 ATTGATTAACATTTAAACATGAAGAATTAAGSTATCTGTTTATGTTAGGATGAG 4604
QY 1010 atagatcttgaggctttatcttgaagatagaatgcaatggaatacttctatgggt 1069
Db 4603 ATGATCTTGAGGCTTTTATTGTAAGATATGAATGAGCATGGAATATGCTTATCTAGGGGT 4544
QY 1070 acttaagaattaaagggtgggtgatttaatttgctcaatctactatcttctaat 1129
Db 4543 ACTTTAAGCTTTAAAGGGGTGGCTGATTTATTTGCTCAATCTCTATATATCTTAAT 4484
QY 1130 atataccagaattataaaagacagatgtccaaattatctctctgcttcaagtaaatg 1189
Db 4483 ATATCATCCAGATTTTAAAGACGATGTCCTCAATATTCCTTCTGCTTCAATTAATG 4424
QY 1190 agatttaaggcttttagagtttcaaaagtcacacagttcaaaagttaaaattcattctt 1249
Db 4423 AGATTTAAGGCTTTTAGAGTTTCAAGTCCACAGTTCAAAAGACAAATTTTCATCTTT 4364
QY 1250 taagtaagaattaaagagttctttatgataagctaagttgttcaatagtttctgatt 1309
Db 4363 TATGTTAAAGATTAAAGATTGATTTATGATTAAGTTGATGCTTTTCAMTAGTTCTGATATT 4304
QY 1310 gacagtagtctgtatcttaaaagttatgagactagcgaactgaatccctcgttaatta 1369
Db 4303 GAGAGTAGCTCTTATTTAAAGTTTATGACACTAGCGGAATCTGATCCCTGTAATTA 4244
QY 1370 aagggcaacgaactttaaagaagtttaaagcttagagaanaaaattctatcgctgaa 1429
Db 4243 AAGGCACAGCAAACTTTTAAAGAGTTTAAAGCTTAGAANAATAATTTCTATGCCGAA 4184
QY 1430 ggtcttttccaaacttggagaaagaattgagagtgaaaaacctgaagatctatccg 1489
Db 4183 GGCTCTTTCNAACCTTTGTAGAAAAGATTGAAGAGTGAACCTGAAAGATCATCTCCG 4124
QY 1490 aaaaattagtttaattaatatcgtaaagctaacctaaaggttgcctteroyssnaca 1549
Db 4123 AAAAATTAGTTTAATTAATATGTAAGCTACTTAAAGGTTGCTT-----ACA 4072
QY 1550 tatltaaatatagaagaatagatatacgttaagaaattagatttggaaaaatgaagacttta 1609
Db 4071 TATTTAAATATATAGAAATAGTATATGGAATATTGATATTGGAATATGAAGGCTTTTA 4012
QY 1610 ggggttttttttgaagaagctcaaatctgtatgatatccttgaagaagaattat 1663
Db 4011 GGAGTTTTTTTTTGAAGAAAGCTCAAAATCTGTAGATATCTTGAAGAGAAATATT 3958

```

## RESULT 3

Q38583

ID Q38583 standard; DNA; 1140 BP.

AC Q38583; 16-JUL-1993 (first entry)

DE 43 kD endoflagellum sheath protein coding sequence.

KW bacteriophage; 43 kD; vaccine; infection; swine dysentery; ss.

OS Treponema hyodysenteriae.

FH Key Location/Qualifiers

FT cds 103..159

FT signal\_peptide 160..1062

FT mat\_peptide 160..1062

```

FT -35_signal /*tag= d
FT 27..32
FT /*tag= e
FT -10_signal 50..55
FT /*tag= f
FT 86..91
FT rbs /*tag= g
FT EP-534526-A.
FT PD 31-MAR-1993.
FT PF 14-SEP-1992: 202796.
FT PR 25-SEP-1991: EP-202478.
FT PR 24-JUL-1992: EP-202273.
FT PA (DUN) DDPHAR INT RES. BV.
FT PI Koopman MBH, Kusters JG;
FT WPI: 93-102665/13.
FT DR P-PSDB: R33279.
FT PT Vaccine to protect pigs against swine dysentery - comprises
FT Treponema hyodysenteriae endo-flagellum sheath protein, applied
FT orally or intranasally
FT PS Claim 2: Page 21-22: 34pp; English.
CC This sequence encodes the endoflagellum sheath protein of T.
CC hyodysenteriae. The endoflagellum consists of at least four
CC proteins, this protein forms the sheath of the flagellum and three
CC proteins, of molecular weights 37, 34 and 32 kD, make up its core.
CC Antibodies raised against the sheath protein have been shown to be
CC bacteriostatic for T. hyodysenteriae. The 43 kD sheath protein can
CC be used in the production of a vaccine against infections such as
CC swine dysentery.
SQ Sequence 1140 BP; 390 A; 151 C; 234 G; 365 T;

```

Query Match 6.4%; Score 107; DB 1; Length 1140;  
 Best Local Similarity 49.3%; Pred. No. 3; le-10;  
 Matches 368; Conservative 0; Mismatches 370; Indels 9; Gaps 3;

```

QY 581 ttatcttagatttgcgagcttgcgaagatgccaaagttcaactagattgattcaca 640
Db 136 TTTATGCTGATCTGCTTACGCTTTAAACAACTTGTGATTTGCTTAAACA 195
QY 641 aattatgtgatattatcatcttcggcgctctctgattgttaagccggaagataagtt 700
Db 196 GGTAAATGCGATTAACCTTAACAGCTGGAAGAGTGAATCAAAATGAACTTCCAGTTGA 255
QY 701 gtatgcttgggataaattgagagctttactactccttcgcaagttgcagct 760
Db 256 GAAATCTTTATATATATTAACCTGAGTAGTAGTTGTTAAATGATCTGCTGATTAACAG 315
QY 761 tacgttaaaatcaagttgttcgcccgtgttcttaagaagttgagtcacaaagttacga 820
Db 316 AATCGCAGAAATCTTATATGTTACTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 375
QY 821 ggtgatactatttggggtaagaagtttgcctcaagctatctc---aatcatgct 877
Db 376 GCAGGTAAAGTCTTCTGTTAAGAGTAAATTTCCCTTATGCAAGCTTGAACAGTTATGCT 435
QY 878 atgattatgcccactttaaatcttcttlatctcgggaaagtgcacaaattctt- 937
Db 436 TTATGTAATAACATATATGAACCTTGAATGTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 495
QY 937 --agcnaagttcttatgataacataaaccataagaagaatgaagttatctgttat 994
Db 496 GAGGTAAAGGTTTATATGATTAACGTTGGCGAAATCAATATATATGTTCTTGCTTAAAT 555
QY 995 agtttggtatgagatagatcttgcggtttatttgaagatagaatgcatggaatat 1054
Db 556 GCAGCTAACTATTTATTTGTTATTTCTGTAACCTTCAAAATGAATTTGCTGAATTAATA 615
QY 1055 gcttattcattggttactttaaagtttaaagggtggcggatttaatttggtcaattcct 1114
Db 616 TCTTATCAATGGGTACTGTTTACTTCAACGTTGGAACAGTAAAGTGAAGGGAACAGACA 675
QY 1115 aactatattcctataatcatcacaagaattatataagaagatgttccaattatctcctt 1174

```





FT	/tag=	l
FT	/label=	Allele II.
FT		3096
FT	/tag=	m
FT	/label=	Allele II.
FT		3140
FT	/tag=	n
FT	/label=	Allele II.
FT		3148
FT	/tag=	O
FT	/label=	Allele II.
FT		3176
FT	/tag=	P
FT	/label=	Allele II.
FT		3182
FT	/tag=	q
FT	/label=	Allele II.
FT		3098..3136
FT	/tag=	r
FT	/note=	"Absent in allele II."
FT		3091..3120
FT	/tag=	s
FT	/label=	Complementary to Probe A.
FT	/note=	"Allele I."
FT		3083..3151
FT	/tag=	t
FT	/label=	Complementary to Probe B.
FT	/note=	"Allele II, skips 3098..3136."
FT		485..2526
FT	/tag=	u
FT	/label=	Regulatory sequence.
PN		M09001549-A.
PD		22-FEB-1990.
PF		10-AUG-1989; 003422.
PR		12-AUG-1988; US-231771.
PA		(DART-) Dartmouth College.
PI		Inselburg JW, Bizik DJ;
DR		WPI: 90-063508/11.
PT		P-PSDB: R05526.
PT		DNA encoding SERA antigen of Plasmodium falciparum -
PT		also DNA encoding signal and regulatory sequences of SERA gene,
PT		for diagnosis, prevention and treatment of malaria.
PS		Disclosure: P: English.
CC		Sequence allows for production of antigenic malarial proteins, useful in
CC		diagnosis, prevention and treatment of the disease. The sequence may be
CC		used in an expression vector, and signal and regulatory sequences may be
CC		used to stimulate production of other products.
SQ		Sequence 6124 BP; 2567 A; 602 C; 801 G; 2137 T; 17 Others;

Query Match 4.0%; Score 67; DB 1; Length 6124;  
Best Local Similarity 47.8%; Pred.No. 0.0016;  
Matches 225; Conservative 0; Mismatches 245; Indels 1; Gaps 1

Db	549	TACACATGATGACAATAAAATGAATGTGATTCTGGTATGATATAAAATTTTATGAAATATA 608
Qy	65	aagatatttaagaagatgaaacaacttgaagaataattagaacctttagttaaagattt 124
Db	609	TAAATTAATGAATATCATATTTGAAAATTTTATACAGATTGTTTCCTTATTTATTTAT 668
Qy	125	tctttaggaaagcaatgaggggttcttccttttggtttttatttaagtgattcaagaatttt 184
Db	669	TTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 728
Qy	185	a-tattcctttttaagaatgaggtttattatagtaaatctttccaaagaattttaagatt 243
Db	729	ATTAACTAGACTTCTTAGTTAAAACTATATATATGTAATTTCTTTTATTTATTAATTTCTTAT 788
Qy	244	ccttaataatggtgatattaatgaatcctttgtagtgcttaagtgcaattccttttgtcatgt 303
Db	789	TTTGCTTATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 848

[illegible]

```

Oy 231 gaattttagatctcttataataatggtgattataatgaatcttttgatgcttaaggtcaat 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 ACAATATGAAAAACAGAGAAATATATGACTTATTTGATATATGAAATATTAAGTTATAT 276
Oy 291 cttttgctatgctttaataaacaatgagcgtttaactatcccgtagaataaagaag 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 AATATTTTAAAGACCTTAAATTTTGTGAATATGTTTGATTTATTTATTTCTTTATGT 216
Oy 351 attatattcttcttgaagggtgatcttaaggagcaaggttgataaatttaac 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 TATATATATGATTTTAAATATATATATATATTTATTTTATTTTATTTTATTTTGAC 156
Oy 411 taataaaataaataaacaagaaattttaaagaattatataagaaggtggtt 470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 AATCTAATTTAATTTTAAATTAATAAATTTATATATATATATATATATTAATTTAAT 96
Oy 471 acatgaagaagcaagctaaagtatttttttttttttttttttttttttttttttt 526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 AAATATATATATATATTTTAAATTTATTTTATTTTATTTTATTTTATTTTATTTT 40

```

## RESULT 7

Q22999/c  
ID Q22999 standard; DNA: 3975 BP.

AC Q22999: 15-JUL-1992 (first entry)

DE SERP gene.

KW Antigen: PHS164-L; SERP; malaria; pOmpA-5; pOmpA-7; ss.

OS Plasmodium falciparum.

FH Key Location/Qualifiers

FT cds 302..3708

FT /tag= a

FT /product= SERP\_antigen

FT exon 302..335

FT /tag= b

FT intron 336..487

FT /tag= c

FT exon 488..1279

FT /tag= d

FT intron 1280..1454

FT /tag= e

FT exon 1455..1595

FT /tag= f

FT intron 1596..1720

FT /tag= g

FT exon 1721..3705

FT /tag= h

FT misc-feature 2077..3430

FT /tag= i

FT /note= "1.3 kb DNA fragment isolated by KpnI and

FT PstI digestion (construction of pOmpA-5) or

FT amplified by PCR (the amplified fragment used

FT for the construction of pOmpA-7)"

FT repeat\_region 520..663

FT /tag= j

FT repeat\_region 1000..1113

FT /tag= k

PN EP-474891-A.

PD 18-MAR-1992.

PF 08-SEP-1990; 117344.

PA (BEHM) BEHRINGERWERKE AG.

PI Schorl J, Knapp B, Hundt E, Kupper H, Amann E;

DR WPI: 92-089689/12.

DR P-PSDB: R22247.

PT Recombinant vector encoding hybrid OmpA protein of Salmonella -

for prodn. of orally-administered malaria vaccines

PS Disclosure: Fig 8; 20pp: English.

CC To construct pOmpA-5, pUC18-SERP which carries the complete SERP gene

CC on a 5.8 kb XbaI fragment (Knapp et al., Mol. Biochem. Parasitol. 32

CC (1989), 73-84) was digested with KpnI and PstI. A 1.3 kb DNA fragment

CC (see features) was isolated and ligated into the respective sites of

CC pSK vectors. Subcloning of the SERP fragment into pSK was performed

CC in order to create the correct translational reading frame at the 3'

CC end of the SERP coding region. The resulting plasmid pSK-SERP was

CC digested with SmaI and KpnI, the 1.3 kb fragment was isolated and

CC cloned into the respective sites of PHS164-L.

CC pOmpA-5 and pOmpA-7 both encode 451 amino acids of the SERP antigen.

SO Sequence 3975 BP; 1650 A; 393 C; 575 G; 1317 T;

Query Match 3.9%; Score 64.8; DB 1; Length 3975;

Best Local Similarity 46.0%; Pred. No. 0.0037;

Matches 219; Conservative 0; Mismatches 257; Indels 0; Gaps 0;

Oy 51 ttgttgctaaattgaatattgaagatgacacaaattgaaagatattgattcct 110

Db 515 TTTATTAACATTTTGTATTAATATTAACACCTAAATAAATAATGCAATGTTGTGAT 456

Oy 111 ttatgaagaagattctcttttgagaagcaatgaggggttctcttttggttttaagt 170

Db 455 GCGTAAATAATATATATTTATTTATTAATTAATTAATTAATTAATTAATTTGTTTCT 396

Oy 171 gattcaagattttatattctttttaagaatgaggttattatgtaattcttcaaga 230

Db 395 TTTTCTTTCCCTTTTCTTTTCCCTTTTCCCTTTTCTTTTCTTTTCTTTTCTTTTCT 336

Oy 231 gaattttagatctctttaataatggtatataatgaatcctttgagttaaggtcaat 290

Db 335 ACAATATGAAAAACAGAGAAATATATGACTTATTTGATATATATATATTAAGTTATAT 276

Oy 291 cttttgctatgctttaataaacaatgagcgtttaactccttgtaagataaagaag 350

Db 275 AATATTTTAAAGACCTTAAATTTTGTGAATATGTTTGATTTTATTTTCTTTTATGT 216

Oy 351 attatattcttcttgaagggtgatcttaaggagcaaggttgataaatttaac 410

Db 215 TATATATATGATTTTAAATATATATATATATTTATTTTATTTTATTTTATTTTGAC 156

Oy 411 taataaaataaataaacaagaaattttaaagaattatataagaaggtggtt 470

Db 155 AATCTAATTTAATTTTAAATTAATAAATTTATATATATATATATATTAATTTAAT 96

Oy 471 acatgaagaagcaagctaaagtatttttttttttttttttttttttttttttttt 526

Db 95 AAATATATATATATTTTAAATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 40

RESULT 8

N60472/c

ID N60472 standard; DNA: 4590 BP.

AC N60472:

DT 24-AUG-1991 (first entry)

DE Sequence encoding the ring-infected Erythrocyte Surface Antigen

DE (RESA).

KW Malaria vaccine; antigen; epitope; ss.

OS Plasmodium falciparum.

FH Key Location/Qualifiers

FT exon 801..995

FT /tag= a

FT exon 1199..4225

FT /tag= b

PN W08601802-A.

PD 27-MAR-1986.

PF 11-SEP-1985; 006960.

PA 11-SEP-1984; AU-007067.

PR 11-SEP-1984; AU-007066.

PR 10-SEP-1985; AU-047326.

PA (HALL-) HALL INST MED RES.

FT	ca...	/tag=
FT		/number=
FT	exon	1111..2202
FT		/tag=
FT		/number=
FT	repeat_region	1249..2202
FT		/tag=
PN	EP-499834-A.	c
PD	26-AUG-1992.	
PF	27-JAN-1992.	
PR	21-FEB-1991; DE-105348.	
PA	(BEHW ) BEHRINGERMERKE AG.	
P1	Hundt E, Knapp B, Kupper H, Nolte D, Kuepper H;	
DR	WPI: 92-286009/35.	
PT	DNA coding protein GBP 130 h and proteins produced - used for	
PT	vaccines to control malaria	
PS	Claim 1; Page 12-14; 17pp; German.	
CC	Glycophorin binding protein 130h is a P.falciparum blood stage	
CC	antigen 69h homologous with the known GBP130. The GBP130h gene and	
CC	its homologues can be used in vaccines to protect against malaria	
CC	caused by P.falciparum. The coding sequence was isolated using PCR	





```
FT misc_binding 3091..3120
FT /tag= s
FT /label=Complementary to Probe A.
FT /note="Allele I."
FT misc_binding 3082..3151
FT /tag= t
FT /label=Complementary to Probe B.
FT /note="Allele II, skips 3098..3136."
FT misc_signal 485..2526
FT /tag= u
FT /label=Regulatory sequence.

W09001549-A.
PD 22-FEB-1990.
PF 10-AUG-1989; 003422.
PR 12-AUG-1988; US-231771.
PA (DART-) Dartmouth College.
PI Inselburg JW, Bzik DJ.
DR WPI: 90-083508/11.
P-PSDB: R05326.
DR DNA encoding SERA antigen of Plasmodium falciparum.
PT also DNA encoding signal and regulatory sequences of SERA gene,
PS for diagnosis, prevention and treatment of malaria.
CC Disclosure: p: English.
CC Sequence allows for production of antigenic malarial proteins, useful in
CC diagnosis, prevention and treatment of the disease. The sequence may be
CC used in an expression vector, and signal and regulatory sequences may be
CC used to stimulate production of other products.
SQ Sequence 6124 BP; 2567 A; 602 C; 801 G; 2137 T; 17 Others;
```

```
Query Match 3.4%; Score 56.8; DB 1; Length 6124;
Best Local Similarity 46.8%; Pred. No. 0.081;
Matches 178; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

OY 147 tttcttttggtttttaaagatgcatcaagattttatattctttttaaagaatgga 206
DB 2524 TTAGATATTGTTGTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTCT 2465

OY 207 gttatcatgtaaatcttcaagagaatttctatgattcttcaataatggtattata 266
DB 2464 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 2405

OY 267 gaactcttgatgcttaagatcactcttctgctatgcttcaataaagaatgcgctt 326
DB 2404 TGAATATGAAATTAAGTATATATATATATTAAGAACTTAATTTTGTGAATATGTTT 2345

OY 327 aactctccggaagataaagaattatattcttctgtaaggggtgatacctaagaag 386
DB 2344 GATTTATTTATTTCTTTATGTTATATATATATATGATTTTATATATATATATTT 2285

OY 387 caaagtgtaataaatttactaataaataaataaataaagaacgaattataa 446
DB 2284 ATTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 2225

OY 447 gattatataataagagtggttctacatgaaagaagaagctaaagctattcttttt 506
DB 2224 TATATATATATATATATATATATATATATATATATATATATATATATATAT 2165

OY 507 tatataccacgtctctttt 526
DB 2164 TTTTATTTATTTATTTT 2145
```

```
RESULT 14
T31530
ID T31530 standard; cDNA; 605 BP.
AC T31530.
DT 15-SEP-1996 (first entry)
DE Human 3' apolipoprotein B SAR element clone Rh32.
KW Erythropoietin; EPO; anaemia; gene therapy; vector;
KW scaffold attachment region; SAR element; apolipoprotein B;
OS transgenic animal; ss.
FT Homo sapiens.
```

```
PN W09619573-A1.
PD 27-JUN-1996.
PF 18-DEC-1995; CA0696.
PR 19-DEC-1994; US-358918.
PA (CANG-) CANGENE CORP.
PI Delcove G.
DR WPI: 96-309587/31.
PT Recombinant DNA molecule expressing mammalian erythropoietin.
PT useful to transform cell lines, and for gene therapy, e.g. of
PT anaemia and other red blood cell disorders
PS Claim 7; Page 59-60; 84pp; English.
CC Human apolipoprotein B (apoB) scaffold attachment region (SAR)
CC element clones Rh32 (T31530) and Rh10 (T31531) respectively carry
CC the 3' human apoB SAR element and the distal 1212 bp 5' human apoB
CC SAR element and 1317 bp proximal sequence. These SAR elements
CC co-map with the boundaries of the human apoB gene chromatin
CC domain. A novel recombinant DNA molecule adapted for transfection
CC of a host cell comprises an erythropoietin (EPO) cDNA (T31529) or
CC genomic clone (T31532) operably linked to an expression control
CC sequence and to the 5' and 3' SAR elements. The SAR elements
CC increase expression of the recombinant EPO in stable, long-term
CC mammalian cell cultures.
SQ Sequence 605 BP; 278 A; 18 C; 35 G; 274 T;
```

```
Query Match 3.4%; Score 56.8; DB 1; Length 605;
Best Local Similarity 47.5%; Pred. No. 0.073;
Matches 205; Conservative 0; Mismatches 222; Indels 5; Gaps 1;

OY 79 atgaacaaattgaagataattagatctttagtaaaagattctttagaagca 138
DB 44 ATTAATAATTTTAAATTAATAATTAATAATTAATAATTAATAATTAATAATTTTATA 103

OY 139 atgaggggttctcttggttttattagtgatgcatcaagatttcttataatcttttaa 198
DB 104 ATTAATAATTTTAAATTAATAATTTTAAATTAATAATTTTAAATTAATAATTTTATA 163

OY 199 agaatggaattatattgtaaatcttcaagagaatttctatgattcttcaataatggtg 258
DB 164 ATTAATAATTTTAAATTAATAATTTTAAATTAATAATTTTAAATTAATAATTTTATA 223

OY 259 attataatgaactcttgatgcttaagatcactcttctgctatgcttcaataaaga 318
DB 224 ATTAATAATTTTAAATTAATAATTTTAAATTAATAATTTTAAATTAATAATTTTATA 278

OY 319 tgcgcttaactcctgtaagaataaagaattatattcttctgtaaggggtatct 378
DB 279 TTATATATTTACATTTTATTTTAAATGTTTAAATTTTACATTTTATTTTAAATGCT 338

OY 379 taaaggagcaaatgtaataatttactaataaataaataaataaagaacgaat 438
DB 339 TTATATATTTACATTTTATTTTAAATGTTTAAATTTTACATTTTATTTTAAATGCT 398

OY 439 ttataaagaattatataataaggggtggttctacatgaaagaagaagctaaagattct 498
DB 399 TTATATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAT 458

OY 499 atttttttatt 510
DB 459 AATTAAGTATT 470
```

```
RESULT 15
O62924/C
ID O62924 standard; cDNA; 8920 BP.
AC O62924.
DT 06-DEC-1994 (first entry)
DE Carbamoyl-phosphate-synthetase II.
KW Carbamoyl-phosphate-synthetase II; CPSII; pscpsII gene;
KW malaria; ss.
OS Plasmodium falciparum.
FT Key
FT cds Location/Qualifiers
FT cds 1226..8401
```







GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 16:18:14 ; Search time 767.16 seconds  
(without alignments)  
4275.931 Million cell updates/sec

Title: US-09-004-395-1

Perfect score: 1663  
Sequence: 1 atgataatcttttttcaaaa.....ataccttgagagaataatt 1663

Scoring table: IDENTITY\_NUC  
Searched: 2546578 seqs, 98626752 residues

Database :

EST: \*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	68.6	4.1	3707	39	C83838	C83838 C83838 Dict
2	63.6	3.8	700	45	AU006427	AU006427 AU006427
3	62.4	3.8	717	45	AU000694	AU000694 AU000694
C 4	61.4	3.7	776	47	A1525944	A1525944 DU145-1.G
5	60.8	3.7	688	45	AU001063	AU001063 AU001063
6	59.6	3.6	593	40	AA949610	AA949610 LD29568.5
7	59.2	3.6	578	40	C90014	C90014 C90014 Dict
8	59	3.5	736	45	AU001091	AU001091 AU001091
C 9	58.2	3.5	315	25	N98113	N98113 2203C3 czep
10	58	3.5	564	38	AA536356	AA536356 LD16932.5
C 11	58	3.5	338	40	C90212	C90212 C90212 Dict
12	58	3.5	645	44	A1292425	A1292425 GH15320.5
13	58	3.5	671	44	A1296407	A1296407 LP10278.5
14	58	3.5	739	47	A1514074	A1514074 GH27130.5
15	58	3.5	708	47	A1517005	A1517005 GH27625.5
16	57.6	3.5	622	47	A1530997	A1530997 SD01824.5
C 17	57.6	3.5	373	25	N98046	N98046 2110C3 czep
18	57.6	3.5	487	44	AU053328	AU053328 AU053328
C 19	57.2	3.4	652	44	A1257829	A1257829 LP06315.5
20	57.2	3.4	711	49	A158398	A158398 AEMTBK45
21	57	3.4	684	44	A1294319	A1294319 LP07659.5
C 22	56.4	3.4	700	48	A1557583	A1557583 PC2.1-1.D
23	56.4	3.4	578	33	AA441179	AA441179 LD16080.5
C 24	56.4	3.4	619	47	A1530991	A1530991 SD01814.5
C 25	56.2	3.4	573	45	AU001034	AU001034 AU001034
26	56.2	3.4	735	49	A1550096	A1550096 AEMTBK27
C 27	55.8	3.4	676	35	AA550485	AA550485 1639m3 gm
C 28	55.6	3.3	450	40	C92473	C92473 C92473 Dict
29	55.4	3.3	547	46	A1456313	A1456313 LD36224.5
30	55.4	3.3	702	49	A1648316	A1648316 AEMTBK90
31	55.2	3.3	593	45	AU002045	AU002045 AU002045
32	55.2	3.3	711	49	A1657490	A1657490 AEMTBK16
33	55	3.3	654	45	AU000948	AU000948 AU000948
34	54.8	3.3	849	45	AU005735	AA439969 LD15031.5
C 35	54.8	3.3	538	51	A1723410	A1723410 hcgl842.T
36	54.8	3.3	538	51	A1723411	A1723411 hcgl842.T
37	54.8	3.3	705	43	A1238819	A1238819 GH14939.5
38	54.6	3.3	636	44	A1293890	A1293890 LP07144.5
39	54.6	3.3	690	49	A1619044	A1619044 AEMTBK38
40	54.6	3.3	699	49	A1619044	A1619044 AEMTBK38
41	54.6	3.3	723	49	A1650127	A1650127 AEMTBK28
42	54.6	3.3	727	49	A1650129	A1650129 AEMTBK40
43	54.6	3.3	707	49	A1650165	A1650165 AEMTBK49
44	54.6	3.3	724	49	A1658446	A1658446 AEMTBK64
45	54.6	3.3	724	49	A1658446	A1658446 AEMTBK64

#### ALIGNMENTS

RESULT 1  
C83838/c 3707 bp mRNA EST 28-APR-1999  
LOCUS C83838 Dictyostellium discoideum SS (H. Urushihara) Dictyostellium  
DEFINITION C83838 Dictyostellium discoideum CDNA clone SSA355, mRNA sequence.  
ACCESSION C83838  
NID 92706770  
VERSION C83838.1 GI:2706770

KEYWORDS EST.  
SOURCE Dictyostelium discoideum.  
ORGANISM Dictyostelium discoideum.  
REFERENCE Eukaryota: Dictyostelida; Dictyostelium.  
AUTHORS 1 (bases 1 to 3707)  
Mori, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M., Yoshino, R., Mita, B. N., Pl. M., Sato, T., Takemoto, K., Yasukawa, H., Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.  
The Dictyostelium developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development  
JOURNAL DNA Res. 5 (6), 335-340 (1998)  
MEDLINE 93156227  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2153367.

CONTACT: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan  
Email: d402huesakura.cc.tsukuba.ac.jp  
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'  
POLYA-No.

FEATURES  
Source  
1. 3707  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/map="11"  
/clone="SSA355"  
/clone\_lib="Dictyostelium discoideum SS (H.Urushihara)"  
/dev\_stage="slug"  
BASE COUNT 1528 a 442 c 502 g 1235 t  
ORIGIN

Query Match 4.1%; Score 68.6; DB 39; Length 3707;  
Best Local Similarity 46.9%; Pred. No. 0.011;  
Matches 346; Conservative 0; Mismatches 384; Indels 7; Gaps 4;

```

OY 2 tgataatccttttcaaaaggtttttatcttctagcaaggaggttctgctaa 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 779 TTAATTTTTTTTATTTGTAATTTTTTTTAATTAATTAATTAATTTTGT 720
OY 62 tttaagataattcaagagatgacaataattgaagaataattagatcttga 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 719 TTGAATCATTTTTTATTTATTTTATTTGTAATTTTTTTTAAGTAATTA 660
OY 122 tttcttttaaggaagaagaggtttcttcttggttttaattgaagtgat 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 659 AATATATTTTTTGTATTTGAATTTTATTTATTTATTAATTTTGTGTTG 600
OY 182 ttataatccttttaagaagatgattatgtaaatctt-caagagaattat 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 599 TTGTAATTTTGTGTTTGTGATTTGATTTTAAGTTAAATTTGTAATTT 540
OY 241 attcctttaataatggtatataaagaaatcttgatgtaaggtaactttg 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 539 ATTATTTGTTTGTGATTTGTAATTTGTTTGTGATTTGATTTTAA 483
OY 301 tgcctttaataaacaagcgcttaactaccggtgaagataaagaattatc 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 482 TATTTGGATTTTATTTTATTTTATTTTATTTGTTATTTTAATTAATTT 423
OY 361 ttgttgaaggtgatcatcaaggagcaagtgataaatttcaataaataat 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 422 TTTTATTTTGTGTTTATTTTAAATTTTATTTTATTTTATTTGTTAT 363
OY 421 aattaaaaaagcaaaatttataaagaattataaagaagtggttacaaga 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 362 TTTTATTTTATTTTATTTTATTTTATTTTATTTTAAATTAATTAATTT 304
OY 481 gaagcctaagatattcttcttaataaccacgtcttcttgcgaagaagc 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

DB 303 TGAAAATGTTATCAAGAATATTTGTTGATTCACCGCCTTTAGCTAAATATGCAGA 244  
OY 541 tgaattacagaagggtcttaaaagcgcaagcctggagaattagcttaatttcgca 600  
DB 243 CCCGAATATATCTATTTTGTGAAAATGTTATCAAGAATATTTGTTATTCACCGCTC 184  
OY 601 gcttcaagagatccaaagtc--aactagacttgatcttaacaatttgatgata 658  
DB 183 TTTAGCTAAATATGCAGACCCGCTATTGACTTTTATGATACAGGTAGATACACTA 124  
OY 659 tattcggcgcttctgtatgttgaagcggagaagatatggttgagatcttggataat 718  
DB 123 TAGATGCTTATGACAAAGCTCTTAACCGGACGCTGGTGTGTTGACTTTTATGATAC 64  
OY 719 aattgagcgtttact 735  
DB 63 AGGTAGACTGATACACT 47

RESULT 2  
AU006427 700 bp mRNA EST 19-JAN-1999  
LOCUS AU006427 Bombyx mori p50(Daizo) Bombyx mori cDNA clone wv41020,  
DEFINITION mRNA sequence.  
ACCESSION AU006427  
NID 94163811  
VERSION AU006427.1 GI:4163811  
KEYWORDS EST.  
SOURCE domestic silkworm.  
ORGANISM Bombyx mori  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
Ditrysia; Bombycoidea; Bombycidae; Bombyx.  
REFERENCE 1 (bases 1 to 700)  
Mita, K., Morimoto, M., Shimada, T., Okano, K. and Maeda, S.  
Establishment of cDNA database of Bombyx mori  
Unpublished (1999)  
JOURNAL On Jan 17, 1998 this sequence version replaced gi:2043379.  
COMMENT

AUTHORS  
TITLE  
JOURNAL  
ORIGIN

FEATURES  
Source  
1. 700  
/organism="Bombyx mori"  
/strain="p50(Daizo)"  
/db\_xref="taxon:7091"  
/clone="wv41020"  
/clone\_lib="Bombyx mori p50(Daizo)"  
BASE COUNT 294 a 25 c 59 g 322 t  
ORIGIN

Query Match 3.8%; Score 63.6; DB 45; Length 700;  
Best Local Similarity 48.6%; Pred. No. 0.11;  
Matches 240; Conservative 0; Mismatches 244; Indels 10; Gaps 2;

```

OY 50 atttggtctaatttaagatattcaagagatgacaataattgaagaataattagatc 109
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 184 AGTATTTAGTATTAATTAATTTTAAATTAATTTGAATGAATGATTTGCTTTTA 243
OY 110 tttagtaagaattcttcttaaggaagcaagaggttcttcttggttttaataag 169
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 244 AATATATCTAGTTTATTAAGAAAATTTTAATTTTATTAATTAATTAATTAAT 303
OY 170 tgattcaagatttataatcttctttaaagaagaggttattatgtaacttcaag 229
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 304 TAAATTTAATAAATATATTTTATTAATAAATTTAAGGATTAATCTTTAAATTT--AAA 361
OY 230 agaatttaagatcttcttaataatgagtgatataaagaaatcttcttgatgtaag 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 362 ATTATTAATAATTTTATTAATAAAGCAATTTATTAATTTATTTATTTGTTAATAAATTTTA 421

Qy 290 tcttttgcctatgctctttaataaacaatgctcttaactctcctgtaagataaaaa 349

Db 422 TTTTATTTAATAATTAATTTTACTTAATGAAAAATTTAATTTTAAATTTTATTAATAA 481

Qy 350 gattaatactctgtgaagggtgatcctaaaggagcaagtgataaattta 409

Db 482 AATAAATTTTAATTTAATTAATTAATTTAATTAATTAATTAATTAATTAATTAATTA 541

Qy 410 ctaataaaaataattaaaaaacgaataatttataaagaatt-----tatataaagg 461

Db 542 ATAAATATTATTAATTTAATTTAATTAATTAATTAATAAAGAAATCGGCAATATTATTAATTC 601

Qy 462 agtggcttacaagaagaagcaatgaatttatttllttatataccactgttc 521

Db 602 ACTGTTTATATCAAAAACATGCTTTTGGTAATAATTTAAAGCTAATCGCCCACTGATA 661

Qy 522 ttcttgcctcaagag 535

Db 662 TATTTATTAAGGC 675

RESULT 3  
AU000694  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AU000694 717 bp mRNA  
AU000694.Bombyx mori p50(Daizo) Bombyx mori cDNA clone e40863, mRNA  
sequence.  
AU000694  
g4156938  
AU000694.1 GI:4156938  
EST.  
domestic silkworm.  
Bombyx mori  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
Ditrysia; Bombycoidea; Bombycidae; Bombyx.  
Mita,K., Morimoto,M., Shimada,T., Okano,K. and Maeda,S.  
Establishment of cDNA database of Bombyx mori  
unpublished (1999)  
On Oct 6, 1998 this sequence version replaced gi:3709815.

```

Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmitsuharu64@nirs.go.jp
PROJECT "CREST project by JST".

FEATURES
    source
        Location/Qualifiers
            1..717
                /organism="Bombyx mori"
                /strain="p50(Daizo)"
                /db_xref="taxon:7091"
                /clone="e40883"
                /clone_1fb="Bombyx mori p50(Daizo)"

BASE COUNT      304 a      21 c      59 g      33 t
ORIGIN

```

0y	82	aacacaatttgcgaagatcaattagatccttttcgtaaaagaattcttccttaaggaagaacaa	141
Db	257	aattatttggaaataaaaatgattatcgattttaaataatctcagattttatgaaaaaaa	316
0y	142	aagggcttccttccttgcgatttcattcaagatcgaacgaatttatatccttttcaaaga	201
Db	317	tttaatttttttttattgattttttttatttaattttaaataaaaaaaattatttttttaata	376

Oy	202	atgaggttactaaagtcaaatccttcacgaagaatttttgattccctttaacaatgvtatt	261
Dd	377	AAATATTTAAGGATAATCTTTAAATTTAAATTTTAAATTTTAAATTTTAAAAAGAAA	436
Oy	262	ataatgaactcttttgaigttaaggltcaactcttttgtcatgtctttaataaaaacaatgc	321
Dd	437	TTTAATATATTTTAAATGTGTTAAATAATTTTATTTATTAATAAATTTTACTAATGA	496
Oy	332	gcttbaactacatcgtgttaagataaaaaagattatatcttctgttgaagggtgatcttaa	381
Dd	497	AAATTAATTTTTTAAATTTTAAATTTTAAATAAATAAATTTTAAATTTAA-----TAAATTA	552
Oy	362	aggagcaaatgtgataaattaacttttctaataaaaaataaataaacogaaaaattta	441
Dd	553	ATTAAATGATTAATAATAGTAAATTTTAAATTTTAAATAAATAAATTTTAAATTTAAATTA	612
Oy	442	taaaagatttatataataaagggtgtttcacatgaaaaagaagctaaagctatttattc	501
Dd	613	ATTAAAAAAAGAATTCGGCAAAATATTATTACCTGTTTATCATCAAAAACAATGCTTT	672
Oy	502	ttttttattaccactgctcttttgctca	531
Dd	673	TTGTTAATTAATTTAAAGTCTATATGCCCA	702

RESULT	4
A1525944/c	
LOCUS	A1525944
DEFINITION	DUI45-1.G08.r DU-145 Homo sapiens CDNA 5' mRNA EST
ACCESSION	A1525944
NID	g9440079
VERSION	A1525944.1 GI:4440079
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo. Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jurnal yu,P.J. and Hood,L. Prostate Cancer Expression Profiling by cDNA Sequencing Analysis Unpublished (1999) On May 18, 1998 this sequence version replaced g1:3137424.
TITLE	
JOURNAL	
COMMENT	

**Contact:** Guyang Matthew Huang  
 Leroy Hood  
 University of Washington  
 Department of Molecular Biotechnology, Box 357730, University of  
 Washington, Seattle, WA 98195  
 Tel: 5106280100  
 Fax: 5106280108  
 Email: huangsm@yahoo.com.  
 Location/Qualifiers  
     1. .776  
     /organism="Homo sapiens"  
     /db\_xref="taxon:9606"  
     /clone\_lib="DU-145"  
     /note="Organ: Cell line; Vector: pBluescript; Directional  
     cDNA library was constructed using Lambda ZP II kit  
     (Stratagene). mRNA was extracted from a prostate tumor  
     cell line DU-145."  
 BASE COUNT      386 a      62 c      77 g      145 t      106 others  
 ORIGIN

Query Match	3.7%	Score 61.4	DB 47	Length 776
Best Local Similarity	41.3%	Pred. NO. 0.24		
Matches 187	Conservative 0	Mismatches 265	Indels 1	Gaps 1
Oy	5	taactcttttttaaaaaggtttttatatttcattcagctgaagagattgtgtcgaattt	64	
DB	756	TAATAATTTTTTATNTNTTTTTTTTTTTTAAATTTTTTTTTTTTTNNNATAATNTTAT	697	

Db	159	AATTAATTGAAAGAAATGTTAATGCTTTTAAATATATCTAGTTTATAGAAAAAA	218
QY	142	aggggttccttccttgggttttataaagatccaagatttatatccttttaaga	201
Db	219	TTTAATTTTATTAATTAAGATTTTATTAATTAATTAATTAATTAATTTTAAATA	278
QY	202	atggaatttatatgtaaatccttcaagaatttatgcttccttttaataatgtagatc	261
Db	279	AAATAATTTTAAGGATTAATCTTTAAATTAATAATTTATATATTTATTTAAAAAGAAA	338
QY	262	ataatgaatcttttgatagtttaagtgcaatccttttgctatgcttctaataaacaatgc	321
Db	339	TTTATTAATTAATTAATTTATTTGTTAATAATAATTTATTAATTAATAATTTTACATAAGAA	398
QY	322	gctttaactatccggtaagaataaaaaagtatttatctctgttgaagggtgatacttaa	381
Db	399	AAATTTATTTTATTTATTTATTTTATTAATAATAATTTTAAATTTTAA-----TAAATTTAA	454
QY	382	agggagcaagtttgtaataaatttactacaaataatcttaaaaaacagaaatttca	441
Db	455	ATTTAAATGATAAACTTAGTAATAATTTATTAATTAATAATAATTTATTAATTTAATTA	514
QY	442	taaaagatttatataaaggagttgtttacatgaaaggaaagcctaaagattttatt	501
Db	515	ATTATCAACCGGAATTCGGCAATATTTTATATTCACCTTTTATATCAAAAAATATGCTTT	574
QY	502	tttttatatccacgttctttttgtcca	531
Db	575	TTGTTAATTAATTTAAAGTCATACTGGCCCA	604

RESULT 6  
AA949610  
LOCUS  
DEFINITION  
ID29566.5prime ID Drosophila melanogaster embryo POT2 Drosophila melanogaster cDNA clone ID29568 5prime similar to U37541:  
Drosophila melanogaster complete mitochondrial genome, mRNA  
sequence.  
AA949610 593 bp mRNA EST 25-NOV-1998  
G3112219  
AA949610.1 GI:3112219  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscocomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 593)  
Harvey, D., Hong, L., Evans-Hoim, M., Pendleton, J., Su, C.,  
Borkstein, P., Lewis, S., and Rubin, G. M.  
BDGP/HMMI Drosophila EST Project  
Unpublished (1997)  
On Jan 9, 1998 this sequence version replaced gi:785426.

Contact: Harvey, D.  
G. M. Rubin-Molecular and Cell Biology  
University of California Berkeley  
539 ISA, Berkeley, CA 94720-3200, USA  
Fax: 510 643 9947  
Email: [http://www.frutify.org/EST\\_est@frutify.berkeley.edu](http://www.frutify.org/EST_est@frutify.berkeley.edu)  
plate: 295 row: F column: 8  
High quality sequence stop: 494.  
Location/Qualifiers  
1..593  
source

```

/c1one_11b="UD Drosophila melanogaster embryo pot2"
/sex="male and female"
/dev stage="0 to 24 hours mixed stage embryonic"
/lab host="X11 Blue"
/note="Organ: embryo; Vector: pot2; Site:1: EcoRI; Site:2:
XhoI; Sited fractionated cDNAs were directly ligated into
"

```

BASE COUNT 263 a 12 c 50 g 268 t  
 ORIGIN

Query Match 3.6%; Score 59.6; DB 40; Length 593;  
 Best Local Similarity 45.5%; Pred. No. 0.49;  
 Matches 251; Conservative 0; Mismatches 299; Indels 2; Gaps 1;

1 atgataccttcccaaaaaggttttattcttccagcaggaggttgccta 60  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 18 ATTAGAATACATTAATAATTAAAGTTTAGTTTGTAAAGAAAAATATTTTAA 77  
 61 attaaagattcaagagcagcaaaatttgaaagataattagactcttctgaaag 120  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 78 ATATTAAGTATTAAGTAAAGAAATTAATTAATTAATTAATTAATTAATTA 137  
 121 attcttcttaagagcaagaggggttcttcttgggttcttcttaagtgatcaagac 180  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 138 AGAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 197  
 181 ttctatacttcttcttaagagaggttcttcttaagagaggttcttcttaagagagag 240  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 198 ATTTTTCATTTTAAAGAGTAAATTAATTAATTAATTAATTAATTAATTAATTA 257  
 241 attcttcttaagagaggttcttcttaagagaggttcttcttaagagagaggttcttct 300  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 258 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 317  
 301 tgccttcaaaacaaacagcgccttcttcttcttcttcttcttcttcttcttcttct 360  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 318 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 377  
 361 ttctgagaggttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 420  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 378 T--TTTAATTTTATGCGATAGCTATTAATTAATTAATTAATTAATTAATTAATTA 435  
 421 aatlaaaaaacgaatttcttaaaagattctataataagaggttgccttcttcttcttcttct 480  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 436 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 495  
 481 gaaagcctaaagattcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 540  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 496 TGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 555  
 541 tggattagcaga 552  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 556 AGTAATTAATGA 567

RESULT 7  
 LOCUS C90014 578 bp mRNA EST 20-APR-1998  
 DEFINITION C90014 Dictyostellium discoideum SS (H. Urushihara) Dictyostellium  
 accession C90014  
 NID 93059634  
 VERSION C90014.1 GI:3059634  
 KEYWORDS EST  
 SOURCE Dictyostellium discoideum.  
 ORGANISM Dictyostellium discoideum.  
 Eukaryota; Dictyostellida; Dictyostellium.  
 1 (bases 1 to 578)  
 YOSHINO, R., MORIO, T. and TANAKA, Y.  
 TITLE Developmental cDNA in Dictyostellium discoideum  
 JOURNAL unpublished (1997)  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2151094.

Contact: Hideko Urushihara  
 Institute of Biological Sciences  
 University of Tsukuba  
 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan  
 Email: d402hu@tsukuba.ac.jp

FEATURES  
 source Location/Qualifiers  
 1..578  
 /organism="Dictyostellium discoideum"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /map="21q"  
 /clone="SSG892"  
 /clone\_11b="Dictyostellium discoideum SS (H. Urushihara)"  
 /dev\_stage="slug"

BASE COUNT 238 a 40 c 91 g 209 t

Query Match 3.6%; Score 59.2; DB 40; Length 578;  
 Best Local Similarity 49.7%; Pred. No. 0.58;  
 Matches 179; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

80 tgaacaaatttgaaagataattagactcttcttcttcttcttcttcttcttcttcttcttct 139  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 215 TGATTAACCTTAGTTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 274  
 140 tggaggggttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 199  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 275 TTGAATTTTCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 334  
 200 gaatgaggttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 259  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 335 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 394  
 260 ctataatgaactcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 316  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 395 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 454  
 317 aatgcgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 376  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 455 ATTGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 514  
 377 cttaaagagccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 436  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 515 GAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 574

RESULT 8  
 LOCUS AU001091 736 bp mRNA EST 15-JAN-1999  
 DEFINITION AU001091 Bombyx mori p50(Daizo) Bombyx mori cDNA clone fhm0036f,  
 accession AU001091  
 NID 94157335  
 VERSION AU001091.1 GI:4157335  
 KEYWORDS EST.  
 SOURCE domestic silkworm.  
 ORGANISM Bombyx mori.  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;  
 Ditrysia; Bombycoidea; Bombycidae; Bombyx.  
 1 (bases 1 to 736)  
 Mita, K., Moriyama, M., Shimada, T., Okano, K. and Maeda, S.  
 TITLE Establishment of cDNA database of Bombyx mori  
 JOURNAL unpublished (1999)  
 COMMENT On Oct 8, 1998 this sequence version replaced gi:3726544.

Contact: Mita K  
 Genome Research Group  
 National Institute of Radiological Sciences  
 Anagawa 4-9-1, Inage, Chiba 263-8555, Japan  
 Email: kmita@nirs64.nirs.go.jp  
 PROJECT = 'CREST Project by JST'.  
 Location/Qualifiers  
 1..736  
 /organism="Bombyx mori"  
 /strain="p50(Daizo)"  
 /db\_xref="taxon:7091"



```

FEATURES
  source
    Institute of Biological Sciences
    University of Tsukuba
    3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan
    Email: d402nu@tsukuba.cc.tsukuba.ac.jp.
    Location/Qualifiers
      1..338
        /organism="Dictyostelium discoideum"
        /strain="AX4"
        /db_xref="taxon:44689"
        /clone="SSI382"
        /clone_1lb="Dictyostelium discoideum SS (H.Urushihara)"
        /dev_stage="slug"
BASE COUNT      177 a      21 c      7 g     133 t
ORIGIN

```

Query Match	3.5%	Score 58:	DB 40:	Length 338:
Best Local Similarity	49.7%	Pred. No. 1:		
Matches 148:	Conservative	0:	Mismatches 150:	Indels 0:
				Gaps 0:
Oy 147	ttctcccttgggtcttcttaagtgatcaagattttatattcttctttaaagaatgga	206		
Db 329	tt	270		
Oy 207	gtttattatgaaatcttccaaggaatttatgattctttaaagaagtgattataat	266		
Db 269	gtatt	210		
Oy 267	gaattcttggatgttaaggtcaatctcttctgaatctttaaataaacaatgcgctt	326		
Db 209	attattattattattattattattattattattattattattattattattattatt	150		
Oy 327	aactacctgtgtaagataaaaaagattattcttctgttgaagggtgatacttaagag	386		
Db 149	attattattattattattattattattattattattattattattattattattatt	90		
Oy 387	caaatgtataaatttaatttacttaacaaataataaacaagaatatttataa	444		
Db 89	aaactgagagaaaaaaagatggggaaaaaaagaaaaaaagaaatgattttttttaa	32		
RESULT 12				
AI292425				
LOCUS				
DEFINITION	AI292425 645 bp mRNA EST 01-DEC-1998			
	GH15320.5prime GH Drosophila melanogaster head port2 Drosophila			
	melanogaster cDNA clone GH15320 5prime similar to M37275:			
	Drosophila melanogaster mitochondrial DNA with 12 tRNAs and 7			
	genes. CDS: complement(2936..4277); PID:g986261, mRNA sequence.			
ACCESSION	AI292425			
NID	g3941832			
VERSION	AI292425.1			
KEYWORDS	EST.			
SOURCE	fruit fly.			
ORGANISM	Drosophila melanogaster			
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE	1 (bases 1 to 645)			
AUTHORS	Harvey,D., Hong,L., Evans-Hoim,M., Pendleton,J., Su,C.,			
	Brockstein,P., Lewis,S. and Rubin,G.M.			
TITLE	BDGP/HHMI Drosophila EST Project			
JOURNAL	Unpublished (1997)			
COMMENT	On May 8, 1995 this sequence version replaced gi:800275.			
FEATURES				

**Source**

```

source
1. 645
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH15320"
/science_lib="GH Drosophila melanogaster head port2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/note="Organ: head; Vector: port2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
port2. Plasmid cDNA library."
BASE COUNT      217 a      34 C      60 g      334 t
ORIGIN

```

Query Match	3.5%	Score 58	DB 44	Length 645
Best Local Similarity	46.8%	Pred. No. 0.88		
Matches 218; Conservative	0	Mismatches 245	Indels 3	Gaps 1

OY	2	tgaatacctctttccaaaaggcttttatttcattccctcgcaaggattggtctaa	61
Db	92	TTAATATATATTTTATTTATTTATTTATTTAAATATTAATTAATGAAGATTATTT	151
OY	62	cttaagacattaaagatgaacaanaatttgaagataatttagactctttagtaaaaga	121
Db	152	AGAAATAAATATTTTGACATTAGAGTATGTGAAGGGCCCTTAGCTTATCAATTTTGTT	211
OY	122	tttctctttagaagcaatgagggctttctctttgggtttttt---attaagtgtccaag	178
Db	212	TCATTAATTCCTACTCAGGTAAATGATATATTTTCAACTTTTAGAATTAATTAATGTAA	271
OY	179	attttatattcctttttaaagaaatgagattatcatgtaaacttccaaagaaattta	238
Db	272	AAATATATTTTATTTTATTTTAAATTCCTTTTGTTTATTAATTAATTAATTAATGAA	331
OY	239	tgaatccttttaataagtgatataataatgaactcttttgatgagtgcaactcctttgc	298
Db	332	TGGTACAAATTAATATATATTTTATTAATTTATTTTATTTTAAATTAATTTTAA	391
OY	299	catgctcttaataaaacaatgcgcttaactatccctgtagataaaagaattatcat	358
Db	392	ATTATTCATCGAAATTTCTTATTTTATTTAGGTGTGATATATTTTCTTAGATTAATTT	451
OY	359	tcttgttgaagggtgatctctaaaggagcaaatgtgataaatttaataatttaataaaa	418
Db	452	TATTAAGTTTATGCAATTTGTCATTAAATTAATTAATTAATTAATTAATTAATTAATTA	511
OY	419	ataattaaanaagaaaaatttataaaagattatataaaggagt	464
Db	512	ATAAATTAATTAATTAATTTATTTATTAATTAATTAATTAATTTATTTATTT	557

RESULT 13

LOCUS	A1296407	671 bp	mRNA	EST	01-DEC-1998
DEFINITION	<p>LP10278.5p1me LP <i>Drosophila melanogaster</i> larval 01-dec-1998  <i>Drosophila melanogaster</i> cDNA clone LP10278 5p1me similar to  M372275: <i>Drosophila melanogaster</i> mitochondrial DNA with 12 tRNAs and  7 genes. CDS:complement(2336..4277); PID:896261, mRNA sequence.</p>				
ACCESSION	A1296407				
NID	9345814				
VERSION	A1296407.1	GI:3945814			
KEYWORDS	EST,				
SOURCE	fruit fly,				
ORGANISM	<i>Drosophila melanogaster</i>				

**COMMENT**

On Jan 17, 1998 this sequence version replaced gi:2044772.

Contact: Harvey, D.  
G. M. Rubin-Molecular and Cell Biology  
University of California Berkeley  
539 USA, Berkeley, CA 94720-3200, USA.  
Fax: 510 643 9947  
Email: [fruityfly.0rg@EST.este@fruityfly.berkeley.edu](mailto:fruityfly.0rg@EST.este@fruityfly.berkeley.edu)  
Plate: 102 row: G column: 6  
High quality sequence stop: 567.

BASE COUNT	222 a	34 c	66 g	348 t	1 others
ORIGIN					
Query Match	3.5%	Score 58;	DB 44;	Length 671;	
Best Local Similarity	46.88%	Pred. NO. 0.87;	245;	Indels 3;	Gaps 1
Matches 218; Conservative					

QY	2	tgaacatctttccaaaagaagttttattttatcttcacgcagaagagtttcgttaa	61
Db	101	TTAAATATATATTTTAAATATATATTTATTTATTTAAATATATTAATATATGAAAGTTATTT	160
QY	62	tttaagaattaaagagatgacacaaatttgaagaatattgatcttttagttaaaga	121
Db	161	AGAAATATATTTTGGACATTTAGAGTATGTGAAGGGCCCTTAGTGTATCAATTTTAGTT	220
QY	122	tttctcttgggaagcaatgaagggtttctcttttgggtttt---attaaigtatccaag	178
Db	221	TCCTAAATGCTACTCATGGAAATGATATTTTCAATCTTTTGAATATATATATGTTAA	280
QY	179	attcttatctctttttaaagaatgaggttattatgtaaactttcaagagaattta	238
Db	281	AAATATATTTTTTTATTTATTTTATTTTATTCCTTTTGTTTTATTTAATATATATATGAA	340
QY	239	tgaattctttaataaagtgagataataatgaactcttttgatgtaaggtcaactctttgc	298
Db	341	TGGTACAAATATATATTTTATTTATTTATTTATTTATTTAATATATATATATATTA	400
QY	299	tatgtctttaataaacaatgcgctttaacttcctgtgaagataaanaagatttat	358
Db	401	ATTATTTGATCGAAATTTCTTATTTTATTTTATTTTATTTTATTTATTTATTTATTTATTT	460
QY	359	tcttcttgaaggggtcatcttaaaaggcgcaagttgataaatttaactactaaaaa	418
Db	461	TATTAAGTTTATGCAATTTGTTCATTAATATTTATTTAGCTTAAGAAATATTAATTAACATA	520
QY	419	ataattaaaaaagcaaaattttaaaaaagttcatataataagaagtc	464
Db	521	ATAATTAATAAAATTTATTTTATTTAATATATATTTATTTATTTATTT 566	

RESULT 14

LOCUS	739 bp	EST	16-MAR-1999
AI514074	mRNA		
GH27130.5	prime GH <i>Drosophila</i> melanogaster head port <i>Drosophila</i>		
melanogaster	cDNA clone GH27130	5prime similar to M37275:	
	<i>Drosophila</i> melanogaster mitochondrial DNA with 12 tRNAs and 7		
	genes. CDS:complement(2936..4277); PID:g896261, mRNA sequence		



ACCESSION	AF1514074
NID	94418136
VERSION	AF1514074.1
KEYWORDS	GI:4418136
SOURCE	EST.
ORGANISM	fruit fly. <i>Drosophila melanogaster</i>
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (phases 1 to 739)
REFERENCE	Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokslein,P., Lewis,S. and Rubin,G.M. BDGP/HMM Drosophila EST Project Unpublished (1997) On Mar 10, 1998 this sequence version replaced gi:2948081.
TITLE	Contact: Harvey, D. G. M. Rubin-Molecular and Cell Biology University of California Berkeley 539 USA, Berkeley, CA 94720-3200, USA Fax: 510 643 9947 Email: http://www.fruitfly.org/Est, estefruitfly.berkeley.edu
JOURNAL	Plate: 271 row: C column: 6 High quality sequence stop: 655.
COMMENT	Location/Qualifiers .1..739 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="GH27130" /clone.lib="GH Drosophila melanogaster head port2" /sex="male and female" /dev_stage="adult" /lab_host="DH5 - alpha" /note="Organ: head; Vector: pORT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pORT2. Plasmid cDNA library." 1 others
BASE COUNT	240 a 43 c 72 g 383 t
ORIGIN	
Query Match	3.5%; Score 58; DB 47; Length 739;
Best Local Similarity	46.8%; Pred. No. 0.85;
Matches 218:	Conservative 0; Mismatches 245; Indels 3; Gaps 1
OY	2 Tgataatccttttcaaaaaaggcttttaatttcactctagcaaggagtgtgtgciaa 61
Db	93 TTATATTATTTTATTTATTTATTTATTTATTTAAATAATTAATTAAGTAATTTT 152
OY	62 tttaagatatttaagaagatgaacaaaatttgaaagataattagactcttgtaaaga 121
Db	153 AGAATAAATATTTTGACATTGTAGAGTAGTGGAAGGGCCCTTAGGTTTATCAATTTTACTT 212
OY	122 tttctcttaggaagcaatgaagggtttctctcttgsggtttt--attaagatcaa 178
Db	213 TCTATATTTCGTACTCATGTGTAATGATTAATTTTCAATCTTTTGAATTAATTAAGTTTAA 272
OY	179 attttcatatctttttaagaagatggagtcttaactglaaatctttccaagaaattta 238
Db	273 AAAATTAATTTTTTTTATTTATTTTATTTATTCCTTTTGTTTATTAATTAATTAATTGAA 332
OY	239 tgatctcttctaataaaggatcataaagaaccttttgatgbtgaagtcacatctttgc 298
Db	333 TGGTACAATAATTAATATTTTATTTATTTGATTTATTTTATTAATAAATAATTAATTTATTA 392
OY	299 taatctcttaataaacaatgcgccttaactaactcggtgaagataaaaaagatatcat 358
Db	393 ATTATTTGATCGAAATTTCTTATTTTATTTTACGTTGTGATTAATTAATTTATTTGGAATTAATTT 452
OY	359 tcctgttgaagggtgatccttaagaagcaaaattgataaatcaaatcttactaataaa 418
Db	453 TATTAAAGTTTATGAAATTTGTTCAATTATTAATTAATTTAGCTGAGAAAATAATTAATAACATA 512
OY	419 ataattaaaaaacgaaaatttatataaagaattcatatataaaggagt 464

Db	513	ATAATTATTAATAAATTTTATTATTATTAATTAATTAATTTATTTATTTATTT	558				
<hr/>							
RESULT	15						
LOCUS	AJ1517005						
DEFINITION	AJ1517005	708 bp	mRNA	EST	16-MAR-1999		
	GH27625.Sprline GH Drosophila melanogaster head por2 Drosophila melanogaster cDNA clone GH27625 Sprime similar to M37275:						
	Drosophila melanogaster mitochondrial DNA with 12 tRNAs and 7 genes. CDS: complement(2936..4277); PID:g896261, mRNA sequence.						
ACCESSION	AJ1517005						
NID	g4420105						
VERSION	AJ1517005.1	GI:4420105.					
KEYWORDS	EST.						
SOURCE	fruit fly.						
ORGANISM	Drosophila melanogaster						
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;						
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;						
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.						
REFERENCE	1 (bases 1 to 708)						
AUTHORS	Harvey,D., Hong,L., Evans-HoIm,M., Pendleton,J., Su,C., Brokslein,P., Lewis,S. and Rubin,G.M.						
TITLE	BGGP/HMTI Drosophila EST Project						
JOURNAL	Unpublished (1997)						
COMMENT	On Mar 20, 1998 this sequence version replaced gi:2979958.						
<hr/>							
Contact:	Harvey, D. G. M. Rubin-Molecular and Cell Biology University of California Berkeley 539 LSA, Berkeley, CA 94720-3200, USA Fax: 510 643 9947 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu Plate: 276 row: C column: 1 High quality sequence stop: 647.						
FEATURES	location/qualifiers						
Source	1..708						
	/organism="Drosophila melanogaster"						
	/db_xref="taxon:7227"						
	/map="6"						
	/clone="GH27625"						
	/clone_1lb="GH Drosophila melanogaster head por2"						
	/sex="male and female"						
	/dev_stage="adult"						
	/lab_host="DH5 - alpha"						
	/note="Organ: head; Vector: pORF2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pORF2. Plasmid cDNA library."						
BASE COUNT	233 a 223 g 364 t 1 others						
ORIGIN							
	Query Match 3.5%; Score 58; DB 47; Length 708; Best Local Similarity 46.8%; Pred. No. 0.86; Indels 3; Gaps 14; Matches 218; Conservative 0; Mismatches 245;						
Oy	2 tgatatacttttcaaaaaaggttcattcatcttagcaaggagtgcgaac 61						
Db	92 TTAATATATTTTTTATATATATATATATATTAATATATTAATATGAAGCTATTTT 151						
Oy	62 tttaagaatatthaagagatgaaacaatttgaaaagacaacttagaccttagtaaaga 121						
Db	153 ACAATAATATTTTTTGACATTTTAGACTAGTAGGAAGGGCCCTTAGGTTCATCAATTTAGTT 211						
Oy	122 tttcttttaggaagaacatggagggttctcctttgggtttt--attaagtgatcaaag 178						
Db	212 TCATATAAATCGTACATCATGTATATATATTTTCAATCCTTTAGAATTAATATATGTTAA 271						
Oy	179 attcttatatactttttaaagaatcgsgattaatgatctaactccccaagaaattta 238						
Db	272 AAATATATTTTTTTTATATATATTTTAAATTCCTTTTGTATTATATATATATATTTGAA 331						
Oy	239 tgattctttaataagtgatcataaagaacttttgtatgtlaaagtgcaactctttgc 298						

```
Db      332 TGGTACAAATTATTAATTTTATTAGATTATTTTATTATTAATAAATTTTATRA 391
Oy      299 taigctttaataaacaatgcgttaactatcctggaagataaaagattatlat 358
Db      392 ATTATTGATCAGAAATTTCTTATTTTTAGGTGATATATTATCTTATGGAITTAATT 451
Oy      359 tcttgtaagggtgtaacttaaggagcaagttgataaatttaactaataaaa 418
Db      452 TATTAGTTTATGAATTTCTCATTAATATATTAGCTAGAGAAATTAATAAACAATA 511
Oy      419 ataattaaaaacgaattttataaagaattatataagaagagt 464
Db      512 ATAAATTATAAAATTTATTTTATTAATAATATATTATTATT 557
```

Search completed: August 18, 1999, 17:04:38  
Job time: 2784 sec



RESULT 2  
W22133  
ID W22133 standard; Protein; 242 AA.  
AC W22133.  
DE 12-JAN-1998 (first entry)  
KW Treponema pallidum rare outer membrane protein (TROMP-2).  
OS Treponema pallidum; Spirochaetales;  
KW Rate; Outer membrane protein; Treponema pallidum; Spirochaetales;  
OS Treponema pallidum.  
FH Treponema pallidum.  
FT Key  
FT Peptide  
FT Location/Qualifiers  
FT 1..24  
FT /note= "putative signal peptide"  
FT 22..24  
FT /note= "cleaves at the C-terminal side of Ala 24"  
FT 25..242  
FT /note= "putative mature protein"  
FT Protein  
FT W09J27215-A1.  
PD 31-JUL-1997.  
PF 23-JAN-1997; 001302.  
PR 23-JAN-1996; US-599480.  
PA (SLOK ) SLOAN KETTERING INST CANCER RES.  
PA (RESC ) UNIV CALIFORNIA.  
PI Blanco DR, Lovitt MA, Miller JN, Tempst PJ;  
DR WPI; 97-393614/36.  
DR N-PSDB; T76579.  
PT New isolated Treponema pallidum outer membrane protein - used in the  
PT detection and induction of immune response in an animal to  
PT pathogenic Spirochaetales  
PS Claim 1; Fig 11; 77pp; English.  
CC This rare outer membrane protein (OMP) of species Treponema pallidum,  
CC suspects pallidum is encoded by a 765 bp sequence (TROMP2 gene).  
CC This is a precursor TROMP2 protein of 242 amino acids. Treponemes  
CC were extracted from infected rabbit testicles and the suspension  
CC tissue debris. Treponemes were obtained from the uppermost zone  
CC gradient after two high speed centrifugations involving a density  
CC to the treponemal suspension which was incubated at room temperature for  
CC 10 minutes and then centrifuged. The treponemal pellet was resuspended  
CC in ice cold sodium citrate buffer. Incubated with occasional  
CC vortexing to release the outer membrane from the inner membrane  
CC Treponemal protoplasmic cylinders were removed by centrifugation.  
CC The outer membrane in the supernatant was subjected to SDS-PAGE to  
CC obtain the rare OMP. The recombinant expression of the rare OMP's  
CC can be used for diagnostic tests to detect syphilis and for  
CC development of host immunity during syphilis.  
SQ Sequence 242 AA.

Query Match 7.3%; Score 125.5; DB 1; Length 242;  
Best Local Similarity 22.4%; Pred. No. 0.00012;  
Matches 55; Conservative 31; Mismatches 104; Indels 55; Gaps 8;  
QY 48 PSTRLDITLVYDYV--YSGASGIYKPEDMVDLGINNW---SVLLTPSARLQAYYKN 100  
DB 21 PLAABRSKYNVQAYFIDEDGAS-----EDQGL-AMRAAGSKFTIKGPFILKYFEG 70  
QY 101 SYVAAPVAVSESKRRYAGDTILIGRVLFPSYSSOSAMIMP-----PKPIFYGESGNOL 155  
DB 71 MQAARMAGSWSGCKDEARFIVGECKFNQGNWMDLITKGGSDVEIPLRGVSG--- 127  
QY 156 GKGLDINIKTMEIKVSVSLGEYIDLEVLFEEDMNGEYVSGTLKFGADLIMSNN 215  
DB 127 -----FDVWVGAGYQYSLDLVDCRGVITLLIGNIDFGKMLSYVPT 173  
QY 216 YIPNISSRIIKDVPNYPYPLASSKMRKAFRVSKSSSKYKFIYKDLRVLYDKLSYI 275  
DB 174 HIPQRS-----RYLGSAGHLSPVGFRIKRTSPSEVRDDFY-----VFDDQKALA 217  
QY 276 DSDID 280  
DB 218 NMHID 222

RESULT 3  
W81355  
ID W81355 standard; Protein; 1657 AA.  
AC W81355.  
DE 16-APR-1999 (first entry)  
KW A. pleuropneumoniae Apxiv toxin serotype 1 protein sequence.  
KW Apxiv toxin; RTX toxin; attenuated microorganism; subunit vaccine;  
OS A. pleuropneumoniae infection; diagnosis.  
FH Actinobacillus pleuropneumoniae.  
FT Ep-875574-A2.  
PD 04-NOV-1998.  
PF 08-APR-1998; 201115.  
PR 10-APR-1997; EP-201032.  
PA (ALKU ) AKZO NOBEL NV.  
PI Frey J, Segers RPAM;  
DR WPI; 99-027956/03.  
DR N-PSDB; V68471.  
PT Attenuated live Actinobacillus pleuropneumoniae - and a nucleotide  
PT sequence having promoter controlling expression of Apxiv gene, a  
PT sub-unit vaccine containing pure Apxiv toxin, and an attenuated live  
PT vaccine.  
PS Example 1; Page 13-20; 56pp; English.  
CC This sequence represents the Apxiv toxin serotype 1 isolated from  
CC Actinobacillus pleuropneumoniae strain 4074 (serotype 1 reference  
CC strain). The invention relates to an attenuated live Actinobacillus  
CC pleuropneumoniae microorganism producing no functional Apxiv toxin. The  
CC invention also provides a subunit vaccine for protecting an animal  
CC against infection of A. pleuropneumoniae, the vaccine contains pure Apxiv  
CC toxin and a pharmacologically allowable carrier. The pure Apxiv toxin can  
CC be used for the diagnosis based on the presence of an antibody against  
CC Apxiv for distinguishing A. pleuropneumoniae infection from A. suis.  
CC N.B. This sequence was indexed from Ep-875574, which is the first major  
CC country equivalent to JP10290695.  
SQ Sequence 1657 AA.

Query Match 6.4%; Score 110; DB 1; Length 1657;  
Best Local Similarity 21.5%; Pred. No. 0.081;  
Matches 68; Conservative 51; Mismatches 114; Indels 84; Gaps 17;  
QY 22 TDGLAEGSKRAEPGEVLIDFAE---LARDPSTR--LDLTNYVYVYSGASGIYKPEDM 75  
DB 860 TEALFNSTFKQSPENALYDLSEVLFENDPTWKEGILLLSRYID--YAKQGFYE----- 914  
QY 76 VVDLGINWWSVLLPISALQAYVKNYSVA---PAYVSESKRRYAGDTILIGRVLFPSYS 131  
DB 914 -----NW-----AATSNLTARLREAGVIFAESTDLKGD----- 943  
QY 132 OSSAMIMPFFKIPFYSGESGNOL--GKGLDINIKTMEIKVSVSLGEYIDLEVLFEEDM 189  
DB 943 EKNNTILISQDNNLSGAGDLDLGSG-NDTLKGSIGATYIFSKHGCD--IYEDT 999  
QY 190 NGMEYAVSMGTLKRGWADLIMSNNYIPNISSRIIKDVPNYPYPLASSKMRKAFRVSKS 249  
DB 1000 NNDNARIDILTKF-----TDVNY-AEYKFRVMDL-----MLFGHDT 1038  
QY 250 HSKYKKNFIYKDLRVLYDKLSISDSDISESYFK-----VYETSGTESLRKKAHET 305  
DB 1039 DSVYTKFSYSYHD--YQFDLLEFA-DRSITRDELIRKAGHLXGTGND--IKDHADW 1091  
QY 306 KRVYLRKIRISIAEGSF 322  
DB 1092 DSILEGGKNDILKGY 1108

RESULT 4  
W73896  
ID W73896 standard; Protein; 1805 AA.  
AC W73896.  
DE 16-APR-1999 (first entry)  
KW A. pleuropneumoniae Apxiv toxin serotype 1 protein sequence.



PT and other proliferative disorders  
 PS Claim 1; Page 19-24; 42p; Japanese.  
 CC The present sequence represents the human transcriptional regulatory  
 CC factor RING3, which is isolated from testicular cells. RING3 contains  
 CC a testis specific bromodomain (TSB) which is expressed specifically  
 CC in testis tissue and also expressed in certain tumour lines. The  
 CC transgenic cells may be used to express RING3 which is a TSB expression  
 CC protein. The TSB expression product can be used in the treatment of  
 CC cancer and other proliferative disorders, and in screening of compounds  
 CC for ability to bind to it (e.g. for use as drugs by modulation of  
 CC transcriptional regulation). DNA capable of hybridising to RING3  
 CC polynucleotides may be used for construction of probes and primers.  
 SO Sequence 947 AA;

Query Match 6.2%; Score 107; DB 1; Length 947;

Best Local Similarity 18.5%; Pred. No. 0.067;  
 Matches 76; Conservative 80; Mismatches 120; Indels 134; Gaps 20;

QY 1 MKRAKSLIFLLSTVLEFAOETDGLAEGSKRAPEGLVDFEALRDPSSST-RIDLNTNV 59  
 DB 118 MAQALEKLFMOKLSQMPQEQVGVKERIKGTQONIAVSSAKKSSPSATEKVFKEQEI 177  
 QY 60 DIVYSGAGIYKPEDMVDGLNNMSVLTTPSARLQAVK---NSVAPAVVSESKRY 115  
 DB 178 PSVFPKTS--ISPLNVVQASVNSSS---QTAQVYTGKVRKADTTTPATSAVKAASSE-- 231  
 QY 116 AGDTILGAVLFPSSYSSASAMIMPFK-----IPFSGESGNOFLGKGLDNKTKME 168  
 DB 231 -----FSPPTFKS--VALPPIKEMPKNVLP---DSQOQY---NVEIVKYTEQ 272  
 QY 169 IK-----VSYSLG---YEI-----DLEVLFEDMNGMEY- 195  
 DB 273 LRHCSEILKEMAKKHFYAMPFYNPVDVNALGLHNYVAVKPNPDGLGIRKKMNOEYK 332  
 QY 195 -AYSMTLKFKGMDLINSN-----PNYIPNISRIIKDDVPNYPPLASSMKRKAFAKRS 247  
 DB 333 DAYS-----FAADVRLMNCYKYNPDHEVVTMARMLQD-----VF 369  
 QY 248 KSHSSKVNKFIYVDLRLYDKLSVSDIDSESVFVETSGTESLRKLKAH----- 303  
 DB 370 ETHFSKIP--IEPVSMPLCYIKTDIT-----ETGRENTEASSEGNSSD 413  
 QY 303 ----ETFRVTLKREKIS-----IAEGSFQNFVEKIESEKPESSPK 340  
 DB 414 DSEDERVKRLAKLOEQLKAVHQQLVLSQVPPRKLKKKSKKREKKK 463

RESULT 7

ID Y07114 standard; Protein: 947 AA.  
 AC Y07114;  
 DT 02-JUL-1999 (first entry)  
 DE W09904265 Seq. ID No. 685.  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer.  
 OS Homo sapiens.  
 PN W09904265-A2.  
 PD 28-JAN-1999.  
 PF 15-JUL-1998; U14679.  
 PR 22-JUN-1998; US-102322.  
 PR 17-JUL-1997; US-896164.  
 PR 10-OCT-1997; US-061599.  
 PR 10-OCT-1997; US-061765.  
 PR 10-OCT-1997; US-948705.  
 PR 11-OCT-1997; GB-021697.  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 PI Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,  
 PI Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E,  
 PI Tureci O;  
 DR WPI: 99-132448/11.  
 PT New isolated cancer associated nucleic acids and polypeptides -

PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 PS Disclosure; Page 728-730; 787p; English.  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 SO Sequence 947 AA;

Query Match 6.2%; Score 107; DB 1; Length 947;

Best Local Similarity 18.5%; Pred. No. 0.067;  
 Matches 76; Conservative 80; Mismatches 120; Indels 134; Gaps 20;

QY 1 MKRAKSLIFLLSTVLEFAOETDGLAEGSKRAPEGLVDFEALRDPSSST-RIDLNTNV 59  
 DB 118 MAQALEKLFMOKLSQMPQEQVGVKERIKGTQONIAVSSAKKSSPSATEKVFKEQEI 177  
 QY 60 DIVYSGAGIYKPEDMVDGLNNMSVLTTPSARLQAVK---NSVAPAVVSESKRY 115  
 DB 178 PSVFPKTS--ISPLNVVQASVNSSS---QTAQVYTGKVRKADTTTPATSAVKAASSE-- 231  
 QY 116 AGDTILGAVLFPSSYSSASAMIMPFK-----IPFSGESGNOFLGKGLDNKTKME 168  
 DB 231 -----FSPPTFKS--VALPPIKEMPKNVLP---DSQOQY---NVEIVKYTEQ 272  
 QY 169 IK-----VSYSLG---YEI-----DLEVLFEDMNGMEY- 195  
 DB 273 LRHCSEILKEMAKKHFYAMPFYNPVDVNALGLHNYVAVKPNPDGLGIRKKMNOEYK 332  
 QY 195 -AYSMTLKFKGMDLINSN-----PNYIPNISRIIKDDVPNYPPLASSMKRKAFAKRS 247  
 DB 333 DAYS-----FAADVRLMNCYKYNPDHEVVTMARMLQD-----VF 369  
 QY 248 KSHSSKVNKFIYVDLRLYDKLSVSDIDSESVFVETSGTESLRKLKAH----- 303  
 DB 370 ETHFSKIP--IEPVSMPLCYIKTDIT-----ETGRENTEASSEGNSSD 413  
 QY 303 ----ETFRVTLKREKIS-----IAEGSFQNFVEKIESEKPESSPK 340  
 DB 414 DSEDERVKRLAKLOEQLKAVHQQLVLSQVPPRKLKKKSKKREKKK 463

RESULT 8

ID W98762 standard; Protein: 450 AA.  
 AC W98762;  
 DT 31-MAR-1999 (first entry)  
 DE H. pylori GHPD 1097 protein.  
 KW GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KW peptic ulcer disease.  
 OS Helicobacter pylori.  
 PN W09843478-A1.  
 PD 08-OCT-1998.  
 PF 01-APR-1998; U06371.  
 PR 29-JUL-1997; US-902615.  
 PR 01-APR-1997; US-833457.  
 PR 24-JUN-1997; US-881227.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 PI Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;  
 DR WPI: 96-342293/46.  
 DR N-PSDB: X14481.

PT New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases  
 PS Claim 8; Page 1640-1642; 2054pp; English.  
 CC This sequence represents a Helicobacter pylori GHPD protein of the  
 CC invention. The polypeptides can be used for preventing or treating  
 CC Helicobacter infections, and gastroduodenal diseases associated with  
 CC these infections, including acute, chronic, and atrophic gastritis, and  
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
 CC used for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.  
 SQ Sequence 450 AA;

Query Match 5.9%; Score 101.5; DB 1; Length 450;  
 Best Local Similarity 23.5%; Pred. No. 0.074;  
 Matches 63; Conservative 41; Mismatches 95; Indels 69; Gaps 16;

QY 95 QAVYKSNVAPAVYKSSKR--YAGDTILGVRLFPYSYSGSSANIM--PEPKIP---FYS 147  
 DB 71 QAVKKNHRIKPLNESTSKVHLKGDMLDKILKQSYSEKIKMIIIDPPYTKNENFIY 130  
 QY 148 GEGSNQGLGKLDINTMKIKSVSLGCEIDLEVLFEDEMGMEYAKSMGTLKFGWA 207  
 DB 131 GDPSQ-----SNEEVLTLTDYSKEKLDY---IKNLF-----GSKCHSGWL 168  
 QY 208 DLWMSNENYIPNISRRIKDD-----VPNYPLASSKM-----RFAF-----RVS 247  
 DB 169 SFMY--PRL--LAKDLKDDGVFIISIDNECQAKLLCDEIRGEENFVACLKWKKKQ 224  
 QY 248 KSHSKV---KNEIFVKDLRVLYDKLSVSDSDISESVFKYETSGTESLKLKAHE 303  
 DB 225 PSEFSKVAVILEVLYVAKFES-LIDRLGLDNVSDSK---PIINTSNMLSKR----- 274  
 QY 304 TEKRVKLREKIS-IAESFONFEKIE 330  
 DB 274 YFKGIRKSDLNFKSGYQNKTKTTE 301

## RESULT 9

W01037  
 ID W01037 standard; Protein; 419 AA.  
 AC W01037;  
 DT 19-JAN-1997 (first entry)  
 DE Mycoplasma 46-48 kDa protective antigen.  
 KM Antigen: vaccine: mycoplasma pneumonia; swine enzootic pneumonia;  
 OS Mycoplasma hyopneumoniae strain Beaufort.  
 FH Key  
 FT misc\_difference 70  
 FT /note- "codon 70 in the nucleotide sequence is  
 FT a stop codon"  
 FT misc\_difference 101  
 FT /note- "codon 101 in the nucleotide sequence is  
 FT a stop codon"  
 FT misc\_difference 254  
 FT /note- "codon 254 in the nucleotide sequence is  
 FT is a stop codon"  
 PN W09628472-A1.  
 PD 19-SEP-1996.  
 PE 15-MAR-1996; AU0149.  
 PR 16-MAR-1995; AU-001789.  
 PA (UYME ) UNIV MELBOURNE.  
 PI Doughty SM, Lee R, Walker J;  
 DR WPI; 96-433763/43.  
 DR N-PSDB; T38241.  
 PT Putative protective antigens against Mycoplasma - used for the  
 PT detection, prevention or treatment of Mycoplasma infections, esp. M.  
 PT hyopneumoniae in swine  
 PS Disclosure: Fig 7; 43pp; English.  
 CC The gene (T38241) coding for a 48 kDa putative protective antigen  
 CC (W01037) was obd. from a Mycoplasma hyopneumoniae genomic library  
 CC by screening with a probe generated by PCR amplification (see also

CC T13814-16). The antigen had originally been isolated from M.  
 CC hyopneumoniae cells using antibody probes enriched with Mycoplasma-  
 CC specific antibodies. Other protective antigens were also identified  
 CC (see also W01024-31). Protective antigens and antibodies can be  
 CC used in vaccines for preventing or treating mycoplasma infections,  
 CC partic. M. hyopneumoniae infections in swine. They can also be used  
 CC for diagnosis.  
 SQ Sequence 419 AA;

Query Match 5.7%; Score 98.5; DB 1; Length 419;  
 Best Local Similarity 20.7%; Pred. No. 0.13;  
 Matches 81; Conservative 48; Mismatches 118; Indels 145; Gaps 20;

QY 11 FLSTVLFAQETDGL-----AEGRAPPG-----ELVDFAPELA-RDPS-- 50  
 DB 9 FLVSSAIVATSLASITAFVAGCGQTSSGTSOSKROAETLKKVSDSRITLTPDNP 68  
 QY 50 ---STRDLTNYVDYVYSGASGIVKPEDVVDGINNMSVLLTPSARLQAVYKSNVAPA 106  
 DB 69 RMISAQKDII SYDETFEATSTTKND-----AQNWW---LQQANLSPAPRGFTIAP 120  
 QY 107 VMSSEKRYAGDTILGVRLFPYSYSGSSANIMPEPKIPFISGESGNO-----F 154  
 DB 121 --NGSGVGTAVNTIADKGIPIVAYDR---LITGSDKIDWVSPDNEKVGELQSLAAGL 175  
 QY 155 LGR--GLIDNFKMKET-----KVSYSLGCEIDLEVLFEDEMGMEYAK---SMGTLK 202  
 DB 176 LGRKEDGFEDSIDDNELKSHMPOETISFTYIAGS-----ODDNNQSYFGAMAKVYLK 228  
 QY 203 FKGADLWMSNENYIPNISR-----IKDD----- 229  
 DB 229 -----ELMKNSQNKIIDLSPEGENAVYPGMNYGTAGARIQSFLTIKKDPAGNKIKAYG 283  
 QY 229 -----VNYPLAS---SKMFAKRVSK-----SHSKVKNFI-----FY 260  
 DB 284 SKPASIFKGFAPNDGAEQATIKLKLEGFTQIKIFVTRQDYNDKAKFTFKDQDNNTIY 343  
 QY 261 VKD-----LRVLYDKLSVSDSDISE 282  
 DB 344 KPDKRVLGKVAVEVLRVLIANKKNSRSEVENE 375

## RESULT 10

R60101  
 ID R60101 standard; Protein; 713 AA.  
 AC R60101;  
 DT 15-MAR-1995 (first entry)  
 DE Canine zona pellucida CZP2.  
 KM Canine; dog; zona pellucida; zp; CZP2; contraceptive; vaccine;  
 OS Canis familiaris.  
 PN J06189766-A.  
 PD 12-JUL-1994.  
 PE 25-DEC-1992; 359265.  
 PR 25-DEC-1992; JP-359265.  
 PA (TOFU ) TONEN CORP.  
 DR WPI; 94-259553/32.  
 DR N-PSDB; Q70072.  
 PT New DNA sequence encoding canine zona pellucida CZP2 - useful for  
 PT the prodn. of a canine contraceptive vaccine antigen  
 PS Claim 1; Page 8-10; 10pp; Japanese.  
 CC The CZP2 DNA (Q70072) was prepd. by the cloning of CZP2(75-520) -  
 CC Q81700 using the primers given in Q70073-74, CZP2(1-65) - Q81804  
 CC using the primers given in Q70082-83, CZP2(42-103) - Q81803 using  
 CC the primers given in Q70079-81 and CZP2(487-713) - Q81957 using the  
 CC primers given in Q70075-78.  
 SQ Sequence 713 AA;

Query Match 5.7%; Score 97.5; DB 1; Length 713;  
 Best Local Similarity 19.4%; Pred. No. 0.38;  
 Matches 79; Conservative 73; Mismatches 128; Indels 127; Gaps 22;





Db 110 DMVVEFAHDLGIR-----VGSNSLSRLIKNIRIFODANGVDFQERYEAVLADSGMTES 163  
 QY 112 S-----KRYAGDTILGVRVLPFSYSOSSAMIMPFKIPFYSGESGNOFLGKGLIDNITKM 166  
 Db 164 SYNNKIRNALPSTIL-MECLEFPNRAE-----LHIFPYDALAKDVVLGL-LQHRVADI 213  
 QY 167 KEIKVSVYSI-GYEI---DLEVLFEEDNGMEYASMGTLFKG-----WADLIW 211  
 Db 214 VEISSDAVIDSGDISDDELQKLFEE---QYKNSLNFPEKRSADYIIMAEDDLADVIY 269  
 QY 212 SNRNYINISRIKD--DYPNPPLA---SSKMFKAFRYSKSSKVKNFIFVYKVL-- 265  
 Db 270 SDEVDVEIKNSSELHDQRDVLNLFETDKNEALYKAYQSGKSEELVSDAGYIEDIAL 329  
 QY 265 -RVLYDKLSYIDSIDISESEFYKYE---TSGTESLRKLKAHTFPRVL-KLREKIS 316  
 Db 330 NNISKDVLPGVRRNVVFNALNEGEVSEMFNSVGVGHMKVIRKHEITKEDLEKLEKIS 387

RESULT 13  
 ID W81975  
 W81975 standard; Protein; 546 AA.  
 AC W81975;  
 DT 02-MAR-1999 (first entry)  
 DE Ehrlichia sp. W20.1 protein.  
 KM Granulocytic ehrlichia; GE; W20; tick-borne infection; fatal; vaccine;  
 KM Immune response; detection; diagnosis; Ehrlichiosis.  
 OS Ehrlichia sp.  
 PN MO9849312-R2.  
 PF 05-NOV-1998.  
 PF 24-APR-1998; 008264.  
 PR 25-APR-1997; US-044869.  
 PA (AQUIT-) AQUILA BIOPHARMACEUTICALS INC.  
 PI Beltz G, Coughlin RT, Murphy C, Storey J;  
 PI WPI; 99-034663/03.  
 DR N-PSDB; V65139.  
 DR N-PSDB;  
 PT New isolated granulocytic ehrlichia nucleic acids - used to develop  
 PT products for use in vaccines for inhibiting Ehrlichiosis and for use  
 PT in detection and diagnosis  
 PS Claim 1a; Fig 1; 184pp: English.  
 CC This sequence represents the granulocytic ehrlichia (GE) protein W20.1  
 CC which has been isolated from HL60 cells infected with Ehrlichia sp. GE  
 CC is an acute potentially fatal tick borne infection and the proteins  
 CC described in this invention can be used in vaccines to elicit a  
 CC beneficial immune response in an animal to GE. They can be used for  
 CC inhibiting Ehrlichiosis in an animal. The products can also be used for  
 CC detection and diagnosis.  
 SQ Sequence 546 AA;

Query Match 5.7%; Score 97.5; DB 1; Length 546;  
 Best Local Similarity 22.8%; Pred. No. 0.25; Mismatches 101; Indels 75; Gaps 16;  
 Matches 68; Conservative 54;

QY 74 DMVV-----DLGINMSVLLTPSARLQAYVKN-----SYVAPAVKSE 111  
 Db 11 DMVVEFAHDLGIR-----VGSNSLSRLIKNIRIFODANGVDFQERYEAVLADSGMTES 64  
 QY 112 S-----KRYAGDTILGVRVLPFSYSOSSAMIMPFKIPFYSGESGNOFLGKGLIDNITKM 166  
 Db 65 SYNNKIRNALPSTIL-MECLEFPNRAE-----LHIFPYDALAKDVVLGL-LQHRVADI 114  
 QY 167 KEIKVSVYSI-GYEI---DLEVLFEEDNGMEYASMGTLFKG-----WADLIW 211  
 Db 115 VEISSDAVIDSGDISDDELQKLFEE---QYKNSLNFPEKRSADYIIMAEDDLADVIY 170  
 QY 212 SNRNYINISRIKD--DYPNPPLA---SSKMFKAFRYSKSSKVKNFIFVYKVL-- 265  
 Db 171 SDEVDVEIKNSSELHDQRDVLNLFETDKNEALYKAYQSGKSEELVSDAGYIEDIAL 230  
 QY 265 -RVLYDKLSYIDSIDISESEFYKYE---TSGTESLRKLKAHTFPRVL-KLREKIS 316

Db 231 NNISKDVLPGVRRNVVFNALNEGEVSEMFNSVGVGHMKVIRKHEITKEDLEKLEKIS 288  
 RESULT 14  
 ID R21829  
 R21829 standard; Protein; 419 AA.  
 AC R21829;  
 DT 02-NOV-1992 (first entry)  
 DE Sequence of surface antigen 46kd.  
 KM Swine pneumonia; epidemic; diagnosis; therapy.  
 OS Mycoplasma hyopneumoniae.  
 PN EP-475185-A.  
 PD 18-MAR-1992.  
 PR 27-AUG-1991; 114335.  
 PR 27-AUG-1990; JP-224945.  
 PA (NIFL-) NIPPON FLOUR MILLS.  
 PI Seto Y, Futo S, Matsuse S, Matsuo K, Tsuna M;  
 PI WPI; 92-089874/12.  
 DR N-PSDB; Q22042.  
 PT DNA and peptide of mycoplasma hyopneumoniae - useful for  
 PT diagnosis and treatment of swine mycoplasma pneumonia  
 PS disclosure: Page 4-6 and pages 34-36; 45pp: English.  
 CC The inventors claim DNA encoding a surface antigen and primers used  
 CC in a method of diagnosing mycoplasma pneumonia of swine.  
 CC Mycoplasma cells were collected from the culture of M.hp, lysed in a  
 CC buffer containing SDS, followed by the extraction of DNA and  
 CC purification thereof. The DNA of M.hp is cleaved with a restriction  
 CC enzyme HindIII, the resulting fragment is inserted into a plasmid  
 CC pUC19 at the site cleaved with HindIII. This plasmid is transformed  
 CC into E. coli as a host cell, colony hybridization is performed by the  
 CC DNA probe method utilizing EcoRI fragments of pKUM1 and pKUM2 to  
 CC select the bacterial cells containing a plasmid (pKUM126) carrying the  
 CC 46 kd antigen gene. E. coli JM 109 which comprises the plasmid  
 CC pKUM1 or pKUM2 is deposited with PRI under accession No. FERM P-10318  
 CC or P-10319. The base sequence of the 46 kd antigen gene is given in  
 CC Q22042.  
 SQ Sequence 419 AA;

Query Match 5.6%; Score 96.5; DB 1; Length 419;  
 Best Local Similarity 20.9%; Pred. No. 0.21; Mismatches 116; Indels 145; Gaps 21;  
 Matches 82; Conservative 49;

QY 11 FLSTVLFQAETGGL-----AEGSKRAEPG-----ELVDFAEFLA-RDPS-- 50  
 Db 9 FLVSAIYARSASIAFAVAGCGQRESGSTDSKQAEFLKKRVSDSRIRALTDPDN 68  
 QY 50 ---STRLDLTNYVDYYSGASGIVKPEDVVDLGINMSVLLTPSARLQAYVKNVYVAPA 106  
 Db 69 RWISAQKDIISYDETEAATSTITKNOD-----AQNWM---LTQANLSPAPLGFITAPE 120  
 QY 107 VKSESKRYAGDTILGVRVLPFSYSOSSAMIMPFKIPFYSGESGNOFLGKGLIDNITKM 175  
 Db 121 -NGSGVGTAVNTADKGIPIVAYDR---LITGSDKIDWVSFDNGVGLGSLAAGL 175  
 QY 155 LGR--GLIDNITKMKET-----KVSYSLGYEIDLEVLFEEDNGMEYAS--SMGTLK 202  
 Db 176 LGEDEGAFEDSIDDMNELLKMHPOETISFTYIAGS-----QDDNNSQPIYGAMKVLK 228  
 QY 203 FKGMADLIWSNPYIPNISR-----IHKD----- 229  
 Db 229 -----ELMKSQNKIIDLSPEGENAVYVPGMYGTAGRIOSFLTIKDPAGKNIKAVG 283  
 QY 229 -----VNYPLAS---SKMFKAPRVSK-----SHSKYKNFI-----FY 260  
 Db 284 SKPASIFLGLAPNDGAEGATIKLKLEGGFTOKIFVTGODYNDKATFTKDDQNNIT 343  
 QY 261 VKD-----LRVLYDKLSYIDSIDISE 282  
 Db 344 KPDVYLGKVAVEVLRLVIAKKNSRSEVENE 375

RESULT 15



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 18, 1999, 17:39:09 ; Search time 11.18 Seconds

(without alignments)  
1222.033 Million cell updates/sec

Title: US-09-004-395-2

Perfect score: 1719  
Sequence: 1 MKRKASILFLSTVLFAQ.....FQNFVERIESEKPESSPKN 341

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database :

PIR-60:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1701	99.0	344	2 C70183	flagellar filament
2	361.5	21.0	350	2 D71348	flagellar filament
3	331.5	19.3	320	1 A43824	periplasmic flagel
4	202.5	11.8	301	2 A32814	flagellar filament
5	121.5	7.1	242	2 H71297	probable outer mem
6	109	6.3	444	2 I40492	gntR-type transcri
7	108	6.3	580	2 F64385	hypothetical prote
8	106.5	6.2	695	2 G64327	H+-transporting AT
9	105.5	6.1	705	2 A41322	N-acetylmutamoyl-L
10	104.5	6.1	1119	2 B70112	surface-located me
11	104	6.0	499	2 D69735	xylose kinase xy
12	103	6.0	650	2 S14949	dnak-type nuclea
13	102.5	6.0	914	2 JCS574	inter-alpha-trypsi
14	102	5.9	241	2 A11298	probable flagellar
15	101.5	5.9	676	2 A45515	dnak-type nuclea
16	101.5	5.9	622	2 A64494	hypothetical prote
17	101.5	5.9	627	2 B64710	adenine specific D
18	101.5	5.9	613	2 JC6033	mosquitocidal prot
19	101	5.9	1060	2 S06286	major merozoite su
20	101	5.9	2166	2 G70163	hypothetical prote
21	100.5	5.8	318	2 E71240	glycyl-tRNA synthe
22	100	5.8	701	2 H71874	inter-alpha-inhibi
23	99.5	5.8	907	2 S54353	hypothetical prote
24	99.5	5.8	401	2 A69368	dnak-type nuclea
25	98.5	5.7	644	2 S14950	46k surface antige
26	98.5	5.7	419	2 A61653	probable membrane
27	98	5.7	635	2 S61175	hypothetical prote
28	98	5.7	707	2 S57157	microtubule-associ
29	97.5	5.7	721	2 A33319	fucosyltransferase
30	97.5	5.7	425	2 G64567	hypothetical prote
31	97	5.6	564	2 T00135	myosin-like protei
32	96	5.6	1875	2 S81173	inculusion protein
33	95.5	5.6	651	2 S46302	inculusion protein
34	95.5	5.6	634	2 T01408	diacylglycerol amide
35	95.5	5.6	931	2 A49737	transcriptional repa
36	95	5.5	1104	2 S36773	transcriptional repa
37	95	5.5	1125	2 F70177	dnak-type nuclea
38	95	5.5	647	2 S4168	dnak-type nuclea
39	95	5.5	642	2 B36590	dnak-type nuclea

40	95	5.5	1701	2 A54498	major merozoite su
41	94.5	5.5	649	2 S53126	dnak-type nuclea
42	94.5	5.5	680	2 S47718	oligopeptidase A (
43	94.5	5.5	436	2 G71862	alpha-(1,3)-fucosy
44	94	5.5	655	2 S18349	dnak-type nuclea
45	94	5.5	430	2 D70193	hypothetical prote

## ALIGNMENTS

RESULT 1  
C70183  
flagellar filament outer layer protein (flaA) homolog - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C>Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 05-Jun-1998  
C:Accession: C70183  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:9805943  
A:Accession: C70183  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-344 <RUE>  
A:Cross-references: GB:AE001166; GB:AE00783; NID:92688596; PID:92688608; TIGR:BB0668  
A:Experimental source: strain B31

Query Match Best Local Similarity 99.08; Score 1701; DB 2; Length 344;  
Matches 337; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRKASILFLSTVLFAQETDGLAEGSKRAEGELVDFAELARPPSTRDLTNVVD 60  
|||||  
DB 4 MKRKASILFLSTVLFAQETDGLAEGSKRAEGELVDFAELARPPSTRDLTNVVD 63  
QY 61 YVSGASGIVPEDMVNDLGNNNSVLTTPSARLQAYKNSVAVPAVYKSKRRACDTI 120  
|||||  
DB 64 YVSGASGIVPEDMVNDLGNNNSVLTTPSARLQAYKNSVAVPAVYKSKRRACDTI 123  
QY 121 LGVRLFPYSYSSAMTIPPKIPYSGESGNOFLKGLDINTMKIKSVSLGDI 180  
|||||  
DB 124 LGVRLFPYSYSSAMTIPPKIPYSGESGNOFLKGLDINTMKIKSVSLGDI 183  
QY 181 DLEVLFEEDMGMEYAYSGTLKFGWADLIWSNPNYIPNISRIIKDDVPYPLASSKMR 240  
|||||  
DB 184 DLEVLFEEDMGMEYAYSGTLKFGWADLIWSNPNYIPNISRIIKDDVPYPLASSKMR 243  
QY 241 FKAFVRSKSHSKYKKNFTFYKDLRYDKLSVSDSDISESVFKYETSGTESLRKLK 300  
|||||  
DB 244 FKAFVRSKSHSKYKKNFTFYKDLRYDKLSVSDSDISESVFKYETSGTESLRKLK 303  
QY 301 AHEPFKRVLKLRKISIAESSGONFVERIESEKPESSPKN 341  
|||||  
DB 304 AHEPFKRVLKLRKISIAESSGONFVERIESEKPESSPKN 344

RESULT 2  
D71348  
flagellar filament outer layer protein flaA-1 precursor - syphilis spirochete  
N:Alternate names: 37k endoflagellar sheath protein flaA; minor endoflagellar protein  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C>Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 17-Mar-1999  
C:Accession: D71348; A40160; A37187; B27559; A33351  
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G  
rison, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uitterlidge, T.; M  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.



```

QY 1 MKRAKSLIFLLSTVLEFAOETDGLAEGSKRAPELVDFEALARDSPSTRLDTNVD 60
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    1 MKR-----FFAIIAGALFVNGSGAFAEQA-----TIDFSLKLVGE-GNGLHAPTTID 47
Db
QY 61 YVSSASIVPED---WVVDGINNWSVLTSPARLQAYVANSVVAAPVAKSESKRRAG 117
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    48 YSRQAGSA-YSAEDKAANKISLAIPSWELASSQTVENQTLSTVTAAPVQDADARVGG 106
Db
QY 118 DTILGVRVLPSPYS-OSSAMIMPFKIPFY-----SGESGNQFLGKGLINDIKTKM 167
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    107 ETVMGVRHIFPSFGINSFAVIKPFTTIPAYATLGDATONAVAGGDFGVLAKNVYIK 166
QY 168 EIKV 171
    |||
    167 SIQI 170
Db

RESULT 5
H71297
Probable outer membrane protein - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-1999
C:Accession: H71297
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin-
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uteerback, T.; MCD
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:96332770
A:Accession: H71297
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-242 <COL>
A:Cross-references: GB:AE001240; GB:AE000520; NID:g3322952; PID:g3322962
A:Experimental source: strain Nichols
A:Genetics:
A:Gene: TP0663

Query Match 7.1%; Score 121.5; DB 2; Length 242;
Best Local Similarity 22.0%; Pred. No. 0.038;
Matches 54; Conservative 31; Mismatches 105; Indels 55; Gaps 8;

QY 48 PSSTRLDTNVDYV---YSGAGIVKPEDVVDGINNW---SVLTSPARLQAYKN 100
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    21 PLAAKSKVNVQAVFIDPFGAS-----EDQGL-AMRAAGSKFTTKGPIIKTFEG 70
Db
QY 101 SVVAPAVVKSSEKRYAGDTILGVRVLPSPYSOSSAMIMPFKIPFYSGESGNQFL 155
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    71 MPQAVRMAGSMQGDKEARFIVECKEFGNQNMDLIPFKGSDYEIPLRGVSG---- 127
Db
QY 156 GKGLINDIKTKMEIKVSVSLGEYIDLEVLFDPMNGMEVAYSGTLKFGWADLINSNP 215
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    127 -----FDVWVWGAQYQSLPALVBDCTGRVHTLLIGNLDFOGMNLSVSP 173
Db
QY 216 YIPNISRLIKDQVNPVPLASSKMRKAFRVKSHSKYKNFIFFYKDLRVLYDKLSVSI 275
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    174 HTPQS-----RYLGSQHLSTFVGFRITXPSERVDY-----VFFDQFKALA 217
Db
QY 276 DSDID 280
    |||
    218 NMHID 222
Db

RESULT 6
140492
gntR-type transcription regulator homolog yxcD - Bacillus subtilis
N:Alternate names: hypothetical protein 8 (srfa operon)
C:Species: Bacillus subtilis
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 24-Sep-1998
C:Accession: 140492; F69766; S34990
R:Cosmina, P.; Rodriguez, F.; de Ferra, F.; Grandi, G.; Perego, M.; Venema, G.; van Sinder

```

```

Mol. Microbiol. 8, 821-831, 1993
A:Title: Sequence and analysis of the genetic locus responsible for surfactin synthesis
A:Reference number: 140485; MUID:93360813
A:Accession: 140492
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-444 <RES>
A:Cross-references: EMBL:X70356; NID:g396480; PID:g396486
A:Experimental source: ATCC 21332
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber-
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erlington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gal-
lechi, J.; Harwood, C.R.; Henuat, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Togomi, A.; Tosato, V.; Uchida
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033
A:Accession: F69766
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-444 <KRN>
A:Cross-references: GB:699105; GB:699106; GB:AL009126; NID:g2632653; PID:e1182323; PI
A:Experimental source: strain 168
A:Genetics:
A:Gene: yxcD
C:Superfamily: hypothetical protein b1439

Query Match 6.3%; Score 109; DB 2; Length 444;
Best Local Similarity 21.6%; Pred. No. 0.77;
Matches 75; Conservative 42; Mismatches 126; Indels 104; Gaps 15;

QY 29 SKRAPELVDFEALARDSPSTRLDTNVD-----YVSGASGT--- 70
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    73 SKSGQGP-IDFATSADP-----DVEPYDFQHCINKAIDTYKNDLFIYTPGSL 125
Db
QY 70 -----VKPDMVVDGINNWSVLT-----TPSARQAYKN 100
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    126 IRYLKLATQGVFADENHITTSVQQAALSLCAMPPEKKEKAIAPQGHILWEQLE 185
Db
QY 101 SVVAPAVVKSSEKRYAGDTILGVRVLPSPYSOSSAMIMPFKIPFYSGESGNQFLGKGLI 160
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    186 TLGIPALIVKRTTE--GIDIAKVEHLFQTESIKPFYTPRFRHP-----LGCSLS 233
Db
QY 161 DNITKMEIKVSVSLGEYIDLEVLFDPMNGMEVAYSGTLKFGWADLINSNPYIPNI 220
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    224 ERDQK-----ELVRLAAYDVYLVEDDY-----LGDEENKAD-----PLYAYDL 274
Db
QY 221 SSRITKQVNPVPLASSKMRKAFRVKSHSKYKNFIFFYKDLRVLYDKLSVSDSID 280
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    275 SSHV-----YLSFSFTMPPGLRVGAVALPEALTDTFYA-----YKKLN-DIDSMI 321
Db
QY 281 SESYKRYETSGTESIRKLAHETEK-RVLLKREKI-----SIAEGSF 322
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    322 SQAALEIYLGSGMGRHKEKIRDSYKESLALHQAIRHROLGSGRF 368
Db

RESULT 7
F64385
hypothetical protein M70686 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997
C:Accession: F64385
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,

```





QY 296 LRKKAHETFKRVLKREKISIAEGS 321  
Db 409 WLQMQADIFNTRVYIKLENGCPAMGA 434

## RESULT 12

S14949

dnak-type molecular chaperone hsc-1 - tomato  
N:Alternate names: heat shock cognate protein 70

C:Species: Lycopersicon esculentum (tomato)

C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 30-Jan-1998

C:Accession: S14949

R:Lin, T.Y.; Duck, N.B.; Winter, J.; Folk, W.R.  
Plant Mol. Biol. 16, 475-478, 1991

A:Title: Sequences of two hsc 70 cDNAs from Lycopersicon esculentum.

A:Reference number: S14949; MUID:91370894

A:Accession: S14949

A:Molecule type: mRNA

A:Residues: 1-650 &lt;LIN&gt;

A:Cross-references: EMBL:X54029; NID:g19255; PID:g19256

C:Genetics:

A:Gene: hsc-1

C:Function:

A:Description: involved in protein folding and assembling/disassembling of protein comp

C:Superfamily: heat shock protein 70

C:Keywords: ATP; molecular chaperone

## Query Match

Best Local Similarity 6.0%; Score 103; DB 2; Length 650;  
Matches 68; Conservative 57; Mismatches 124; Indels 112; Gaps 13;

QY 43 ELARPPSTRDLTNNVYVYGASGIVKPEDMVVDLGINNMSVLTTPSARLQAVYKNSV 102  
Db 61 QVALNININVFPAKRLIGRRFSDAS--VQEDMKL-----WPKYIPEPGDK----- 105

QY 103 VAPAVV---KSKSKRYAGDTILGVRVLPSPYSOSSAMINPPK----- 143  
Db 105 --PMIVYVYKGEKEFEAAEEI-----SSMVLTKKKEIAEAFGSTVKNAVYT 149

QY 143 IPFYSGESGNFL-GKGLDINIKMKET--KGVYSIGYIDLE-----VLFEDMNGM 192  
Db 150 VPAFYENDSORQATKAGAVISGLNWKIINEPTAAATAYGLDKKAVSAGEKNVLIIDLG 209

QY 193 EYAVSMGTLK-----FKGMADLIMSNPYIYNISRIITKDDVPVYPLASSKM 239  
Db 210 TFDVSLTIEEGIFEVKATAGDTHLGEFDNRMVNHFYHEKRRKKDKITGNPRALRL 269

QY 240 RFRAPRVSKSHSKVKNFI-----FYVKDLRVLYDKLSVSI-----D 276  
Db 270 RFACEFAKRTLTSTAGTTTIEDISLVEGVFYSTIFRAREEINMDLFRCKMEPVEKCLRD 329

QY 277 SIDSISVYKYVETSTELRKLKAHETFKRVLKREKISIAEGSQNVVEIESKPEE 336  
Db 330 AKMDSTVHDVVLVGSTRIPIVQ-----QVAMTFNFKGELCKSLNPDE 374

QY 337 S 337  
Db 375 A 375

## RESULT 13

JC5574

Inter-alpha-trypsin inhibitor heavy chain 1 - golden hamster

C:Species: Mesocricetus auratus (golden hamster)

C>Date: 23-Sep-1997 #sequence\_revision 23-Sep-1997 #text\_change 17-Mar-1999

C:Accession: JC5574; PC4484

R:Nakatani, T.; Suzuki, Y.; Yamamoto, T.; Sinohara, H.

J. Biochem. 122, 71-82, 1997

A:Title: Molecular cloning and sequencing of cDNAs encoding three heavy-chain precursors

sin inhibitor heavy chain family.

A:Reference number: JC5574; MUID:97420688

A:Accession: JC5574  
A:Molecule type: mRNA  
A:Residues: 1-914 <NAK>  
A:Cross-references: DDBJ:D89285; NID:g1694687; PID:d1014635; PID:g1694688  
A:Experimental source: liver

A:Accession: PC4484

A:Molecule type: protein

A:Residues: 387-400;461-475 &lt;NA2&gt;

C:Comment: In the plasma three inter-alpha-trypsin inhibitor heavy chains 1, 2 and 3  
that the complexes play important role for pancreatic cancer.

C:Superfamily: inter-alpha-trypsin inhibitor complex component II

F;247-250,690-890/Disulfide bonds: #status predicted

## Query Match

Best Local Similarity 6.0%; Score 102.5; DB 2; Length 914;  
Matches 66; Conservative 59; Mismatches 119; Indels 101; Gaps 15;

QY 16 VLFAQETDGLAEGSKRAPEGLVDFAEIARDPSTRDLTNNVYVYGASGIVKPED- 75  
Db 296 LVFVIDIGSGMEGQKVKOTKEALKIL-----GDVKKPGDS 330

QY 75 --MYVDGINNMSVLLTPS--ARLQA---YKNSVAPAVYKSESKRYAGDTILGVRVL 126  
Db 331 FDLVFGSRVQSMKSLVPAQANLQAQDFVRRPSLAGAT-----NLNGLLGLLEIL 384

QY 127 -----FPSYQSSAMIMPPEKIPFYSGESGNQFLGKGLDINIKTKRKIVSV-----Y 174  
Db 385 NKAQSGHPELSSPASIL-----IMLTGEP-----TEBETDRSQILKNVRAINAGRPFLY 434

QY 175 SLGEYIDLEVLPEDMNGEYVSMGTLKFKGMADLIMSN-----PNYIPNISRITKD 227  
Db 435 NLGFGHLDLDFNLEWMSYENS-----GMAQRIYEDHDATQOLGFGYQVANPLITD 485

QY 228 DVPNPP---LASSMKRRAE-----RVKSKSHSKYKNTIFVKKDLRVLYDKLSV 273  
Db 486 VELQPPQDSVSLTQHRKQYDGSSEIYVAGRDIADHKILSTK-----ADVRRNGERQEF 539

QY 274 SIDSIDISESVFYKYEYTGSG---TESLRKLKAHETFKRVLKREK 315  
Db 540 KATCLVDEEMKRLRERGHLENHVERLMAYLTIQELLARMKM 584

RESULT 14  
A11298  
probable flagellar filament outer layer protein (fla-2) - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C>Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 17-Mar-1999

C:Accession: A11298

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G  
rison, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; M  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A11298; MUID:98332770

A:Accession: A11298

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Cross-references: GB:AE001240; GB:AE000520; NID:g3322952; PID:g3322963

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0664

## Query Match

Best Local Similarity 5.9%; Score 102; DB 2; Length 241;  
Matches 42; Conservative 17; Mismatches 72; Indels 20; Gaps 7;

QY 142 KIPFYSGSGNQLGKGLDINIKTM-----KEIKVSYSLGYIDLEVLPEDMNGMEYA 195  
Db 99 KVSFTR-RGYNF-----EVSSVKKPLPYEGVAKTVSWWAGRGYHSLLEDEWGORFE 153



QY 196 YSMGTLKFKGWADL-ITSNPNYIPNISRRIKD-DVNPYPLASSKMKRFAFVSKSHSK 253  
||| ||| : : : : : ||| : : |||  
Db 154 LHMGLDFSGMKLMSVAIPQHVHDKGTGIQKSRFPN----QSGLNIVGFRVDCDPLEA 209  
QY 254 VKNFIFFVNDLRY---LYDKLSVSDSDIDS 281  
||| ||| ||| ||| : : : : :  
Db 210 YGNIVYVFDLRYVTDLYVESREPDMDSDA 240

## RESULT 15

A45515  
dnaK-type molecular chaperone - Trypanosoma brucei  
N:Alternate names: heat shock-related protein  
C:Species: Trypanosoma brucei  
C:Date: 14-May-1993 #sequence\_rev150n 14-May-1993 #text\_change 13-Mar-1998  
C:Accession: A45515  
R:Lee, M.G.S.: Polvere, R.I., Van der Ploeg, L.H.T.  
Mol. Biochem. Parasitol. 41, 213-220, 1990  
A:title: Evidence for segmental gene conversion between a cognate hsp70 gene and the tem  
A:reference number: A45515  
A:Accession: A45515  
A:status: preliminary  
A:molecule type: DNA  
A:Residues: 1-676 <LEP>  
A:Cross-references: GB:M32139; NID:g162124; PID:g162125  
C:Function:  
A:Description: Involved in protein folding and assembling/disassembling of protein comp  
C:superfamily: heat shock protein 70  
C:Keywords: ATP; molecular chaperone

Query Match 5.9%; Score 101.5; DB 2; Length 676;  
Best local Similarity 23.2%; Pred. No. 5.1;  
Matches 73; Conservative 44; Mismatches 112; Indels 85; Gaps 16;  
QY 63 YSGASGIVKPEDAVNDG-----INMSVLTPLSARLQAIYKNSYAP 105  
||| ||| ||| : : : ||| : : |||  
Db 3 YEGAIGI-----DGTYYSCVGMQNERVEIANDQGNRTTPS--YVAFVNEEVLVG 52  
QY 106 AVKSESKRYAGDTILGV-RVLPFSYSOS--SAMIMPPRIPIYSGESGN-----QFLG 156  
||| : : : ||| : : : ||| : : : |||  
Db 53 DAKSHAKRSGNGVIFPAKRLIGKRFSDSYQSDMKHMPKVE--EGEKGAVMKVEHLG 110  
QY 157 KGLIDNKTKEIKVSYSLGEIDLEVLFDNMGMEYASMGTLKFKGWADLITSNPNY 216  
||| : : : ||| : : : ||| : : : |||  
Db 111 EGMU-----LQPEQISARVLAY--LKSCAESYLGKQVAKAVTV-----PAY 150  
QY 217 IPNISRRIKD-----DVNPYPLASSKMKRFAFVSKSHSKVKNFIFFVNDLRY 266  
||| : : : ||| : : : ||| : : : |||  
Db 151 FNSQROATRDAGTACLEVLRIINEPTAA--IAYGLKADGKGRNVL----- 199  
QY 267 LYDKLSVSDSDIDS--VEKYYETSGTESL--RKLAHETFKRVKLREKISIAESGF 322  
||| : : : ||| : : : ||| : : : |||  
Db 199 VFDEGGGTFFVSTISVSGVFEVAKATNGDTHLGEDVDALLHEHALDIRNRIGIEQSL 258  
QY 323 -QNFVEKIESEKPE 335  
||| : : : ||| : : : |||  
Db 259 SOKMLSKLRSRCEE 272

Search completed: August 18, 1999, 18:14:05  
Job time: 2096 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 1999, 09:33:06 ; Search time 10.76 Seconds

(without alignments)  
895.863 Million cell updates/sec

Title: US-09-004-395-2

Perfect score: 1719

Sequence: 1 MKRAKSLIFLLSTVLAFAQ.....FONFVKIESEKPESSPKN 341

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database: SwissProt\_37.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	403	23.4	337	1	FLAA_SPIAU	P21982 spirochaeta
2	361.5	21.0	350	1	FLAA_TREPA	P18193 treponema p
3	331.5	19.3	320	1	FLAI_TREH	P32520 treponema h
4	109	6.3	444	1	YCXD_BACSU	O08792 bacillus su
5	108	6.3	580	1	Y686_METJA	O58099 methanococ
6	106.5	6.2	695	1	ATP1_METJA	O27675 methanococ
7	105.5	6.1	705	1	LYTB_BACSU	O02113 bacillus su
8	105	6.1	902	1	ITH1_PIG	O29052 sus scrofa
9	104	6.1	499	1	XYLB_BACSU	P39211 bacillus su
10	103	6.0	650	1	HS71_LYCES	P29052 sus scrofa
11	102.5	6.0	914	1	ITH1_MESAU	P4659 lycopodium
12	101.5	5.9	676	1	HS7C_TRYBB	P97278 mesocricetu
13	101.5	5.9	622	1	YF54_METJA	P20030 trypanosoma
14	101	5.9	1682	1	MSP1_PLAF3	O58949 methanococ
15	99.5	5.8	907	1	ITH1_MOUSE	P19598 plasmodium
16	98.5	5.7	644	1	HS72_LYCES	O61702 mus musculu
17	98.5	5.7	416	1	P46_MYCHY	P27332 lycopodium
18	97.5	5.7	707	1	YJ9C_YEAST	P46192 mycoplasma
19	97.5	5.7	721	1	ENP1_TORCA	P47166 saccharomyc
20	96	5.6	1875	1	MUP1_YEAST	P14400 torpido cal
21	95.5	5.6	1104	1	BUD2_YEAST	O02455 saccharomyc
22	95.5	5.6	931	1	DAP1_YEAST	P33314 saccharomyc
23	95.5	5.6	434	1	ENO_LOLPE	O02654 saccharomyc
24	95.5	5.6	651	1	HS71_ARATH	O02654 loligo peal
25	95	5.5	1125	1	HS74_YEAST	P22953 arabidopsis
26	95	5.5	1125	1	MED_BORBU	P22202 saccharomyc
27	95	5.5	1701	1	MSP1_PLAF	O51566 borrelia bu
28	95	5.5	1024	1	SYIP_STAUF	P13819 plasmodium
29	94.5	5.5	680	1	OPDA_ECOLI	P41368 staphylococ
30	94.5	5.5	635	1	HS70_DAUCA	P27298 escherichia
31	93.5	5.4	494	1	AMY1_SACFI	P26791 daucus caro
32	93.5	5.4	651	1	HS7C_PETHY	P21567 saccharomyc
33	93.5	5.4	406	1	RENI_HUMAN	P09189 petunia hyb
34	93	5.4	1296	1	ASAL_ENTFA	P00797 homo sapien
35	93	5.4	608	1	KU70_HUMAN	P12953 enterococcu
36	93	5.4	341	1	MDHM_BRANA	P12956 homo sapien
37	93	5.4	1701	1	MSP1_PLAFM	O43744 brassica na
38	93	5.4	282	1	RU1A_HUMAN	P08569 plasmodium
39	93	5.4	282	1	RU1A_XENLA	P09012 homo sapien
40	93	5.4	701	1	SYGB_HELPY	P54429 xenopus lae
41	92	5.4	517	1	EUG1_YEAST	P56454 heliobacte
42	92	5.4	679	1	GR78_KLUDIA	P32474 saccharomyc
43	92	5.4	1203	1	MGR5_RAT	P22010 kluyveromyc
						P31424 rattus norv

## ALIGNMENTS

RESULT	1	FLAA_SPIAU	STANDARD	PRT	337 AA	P17744 haemophilus
AC	P21982:					O51068 borrelia bu
DT	01-AUG-1991 (REL. 19, CREATED)					
DT	01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)					
DT	01-AUG-1991 (REL. 19, LAST ANNOTATION UPDATE)					
DE	FLAGELLAR FILAMENT OUTER LAYER PROTEIN PRECURSOR (SHEATH PROTEIN).					
GN	FLAA.					
OS	SPIROCHAETA AURANTIA.					
OC	BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; SPIROCHAETA.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE; 89155480.					
RA	ABRAMS B., GREENBERG E.P.:					
RT	"Cloning and sequence analysis of flaa, a gene encoding a Spirochaeta					
RT	aurantia flagellar filament surface antigen."					
RT	J. BACTERIOL. 171:1692-1697(1989).					
RN	[2]					
RP	REVISIONS.					
RA	GREENBERG E.P.:					
RL	SUBMITTED (JAN-1991) TO EMBL/GENBANK/DBJ DATA BANKS.					
RN	[3]					
RP	SEQUENCE OF 22-49.					
RC	STRAIN-M1;					
RX	MEDLINE; 91123217.					
RA	PARALES J. JR., GREENBERG E.P.:					
RT	"N-terminal amino acid sequences and amino acid compositions of the					
RT	Spirochaeta aurantia flagellar filament polypeptides."					
RT	J. BACTERIOL. 173:1357-1359(1991).					
CC	-1- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.					
CC	-1- SUBUNIT: OUTER LAYER COMPOSED OF REPEATING UNITS OF FLAA AND A					
CC	CORE THAT CONTAINS ONE OR ALL OF FIVE ANTIGENICALLY RELATED					
CC	POLYPEPTIDE.					
CC	-1- SUBCELLULAR LOCATION: PERIPLASMIC FLAGELLUM.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>					
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).					
CC	-----					
DR	EMBL; M24459; G152896; -					
DR	PIR; A32814; A32814.					
KW	FLAGELLA; PERIPLASMIC; SIGNAL.					
FT	SIGNAL					
FT	CHAIN					
FT	22 337					
FT	SEQUENCE					
FT	337 AA; 36857 MW; B6AE70C3 CAC32.					
SO	FLAGELLAR FILAMENT SURFACE ANTIGEN.					
Query Match	23.4%	Score 403;	DB 1;	Length 337;		
Best Local Similarity	30.4%	Pred. No. 2.2e-22;				
Matches 100;	Conservative 67;	Mismatches 124;	Indels 38;	Gaps 9;		
DB	1 MKRAKSLIFLLSTVLAFAQETDGLAGSKRAEGELVLPFAELARPSSRLDLYVD 60					
DB	1 MKR-----FFAIIAGALFVNGSAGFAQA-----TLIDFSKLVGE-GNTGLAAPTID 47					
DB	1 YVYGAGSIVPED---MVVDLGINNSVLLTPSARIQAVYKNSVVAIPAVKSSSKRYAG 117					
DB	48 YSRAGSA-YAEDKAMKISLAPSWIEIASSQVENVNTLSLVTAAPKQDAATGG 106					
DB	118 DTILGVRLPSPYS-QSSAMIMPFKIPY-----SGESGNOFLGKGLDINIKTMK 167					

DB 107 ETWGVRIHFPSEFGINSFAVIRKPEPTIPAYALGDATAQNAVAGQFDFGVAKNVGK 166  
 QY 168 ETKSVYSVSGYEIDLEVFEDNMNGMEYASMGTLKFGWADLIWSNPNTIPNISRIRD 227  
 DB 167 STQINLGNINYLNRLLSDLEDONGDERELVNGVLYNFGDKSLQNNPNPTQTEVRNDDLO- 226  
 QY 228 DVNPFPPLASSKMKRFAFRVSKSHSKVKNFFIYVNDLRYLYKLSVSDSIDSESV--- 285  
 DB 226 IYPLPRSPAPLTKGKIKHRDQSGGDIVSYIKDKIKYIDQAVYDRNSVDVDEDAIWGI 285  
 QY 285 -----FKVYETSGTESIRKIKAHETK 306  
 DB 286 LKREQYRNFEALQKGLQVLRSLSEKK 314

RESULT 2  
 FLAA\_TREPA STANDARD: PRT: 350 AA.  
 ID FLAA\_TREPA  
 AC P18193:  
 DT 01-NOV-1990 (REL. 16, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE FLABELLAR FILAMENT OUTER LAYER PROTEIN PRECURSOR (SHEATH PROTEIN).  
 GN FLAA OR TP0249.  
 OS TREPONEMA PALLIDUM.  
 OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; TREPONEMA.  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 90307197.  
 RA ISANCS R.D., RADOLF J.D.;  
 RT "Expression in Escherichia coli of the 37-kilodalton endoflagellar  
 RT sheath protein of Treponema pallidum by use of the polymerase chain  
 RT reaction and a T7 expression system.";  
 RL INFECT. IMMUN. 58:2025-2034(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NICHOLS;  
 RX MEDLINE: 98332770.  
 RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,  
 RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,  
 RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,  
 RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDIMBARAM M., UTTERBACK T.,  
 RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,  
 RA HATCH B., HORST K., ROBERTS K., MATTHEY L., WEIDMAN J., SMITH H.O.,  
 RA VENTER J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete.";  
 RL SCIENCE 281:375-388(1998).  
 RN [3]  
 RP SEQUENCE OF 30-350 FROM N.A.  
 RC STRAIN-NICHOLS;  
 RX MEDLINE: 90035409.  
 RA ISANCS R.D., HANKE J.H., GUZMAN-VERDIZCO L.-M., NEWPORT G.,  
 RA AGADIAN N., NORCARD M.V., LUKEHART S.A., RADOLF J.D.;  
 RT "Molecular cloning and DNA sequence analysis of the 37-kilodalton  
 RT endoflagellar sheath protein gene of Treponema pallidum.";  
 RL INFECT. IMMUN. 57:3403-3411(1989).  
 CC -1- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.  
 CC -1- SUBUNIT: OUTER LAYER COMPOSED OF REPEATING UNITS OF FLAA AND A  
 CC CORE THAT CONTAINS ONE OR ALL OF FIVE ANTIGENICALLY RELATED  
 CC POLYPEPTIDE.  
 CC -1- SUBCELLULAR LOCATION: PERIPLASMIC FLAGELLUM.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M63142; G155059;  
 DR EMBL: AE001206; G3322518;

DR EMBL: M26525; G155057; -  
 DR TIGR: TP0249; -  
 KW FLAGELLA; PERIPLASMIC; SIGNAL.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 350 FLAGELLAR FILAMENT SURFACE ANTIGEN.  
 SO SEQUENCE 350 AA; 38857 MW; 640702B9 CRC32;

Query Match 21.0%; Score 361.5; DB 1; Length 350;  
 Best Local Similarity 28.5%; Pred. No. 2,4e-19;  
 Matches 100; Conservative 66; Mismatches 120; Indels 65; Gaps 11;

QY 6 KSIL--FELLSTVLEFAQDTGLAEGSKRAPEGEVLDFAEIARDPSSRTLDLTNYVDIVY 63  
 DB 22 ESVLIDFAKLNADIMADSGGNTNHR-----TVLDYASILD-----TSYTD--- 64  
 QY 64 SGASGIVPEDMVYDGLNNMSVLLTPARLQATYKNSVAPAYVSKSRKRYAGDTILGV 123  
 DB 64 -----EOKALMRSSSLAOWEVLNSSARNPVAAASRVIEAPVSEGAKSFADERLVGV 117  
 QY 124 RVLFPSTY-SOSSAMIMPFFKIPFYS-----GEGN-----QFL-GKGLIDN 162  
 DB 118 RVLFPSTMSNNAMKPAFVIPAYVMAQVDDGVQAPTEEEKASGKRGREDIGVYKN 177  
 QY 163 IKTKELRVYSVSLGEYIDLEVFEDNMNGMEYASMGTLKFGWADLIWSNPNTIPNISR 222  
 DB 178 VGLKSIAYNTYGMNYPGLIYVMRDQGEVHRFYMGYLTFDSWKEIYVNNPNYSIDVRS 237  
 QY 223 RIIRKDVNPYPLASSKMKRFAFRVSKSHSKVKNFFIYVNDLRYLYKLSVSDSIDSE 282  
 DB 228 REVR-LYVPYPASTPHVVEEPMYTRDAHAAGGDVGVFKVKKIYDAVASTVARDFADE 296  
 QY 283 SVFKYETSGTESIRKIKAHETKFRVLRKIKISIAESGFONFEKIRSEK 333  
 DB 297 DLW-----GIQARREAE-----RKREVARRGQGVLRITQEK 330

RESULT 3  
 FLAL\_TREHY STANDARD: PRT: 320 AA.  
 ID FLAL\_TREHY  
 AC P32520; P80157;  
 DT 01-OCT-1993 (REL. 27, CREATED)  
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN FLAAL PRECURSOR (44 KD SHEATH  
 DE PROTEIN).  
 GN FLAAL OR FLAA.  
 OS TREPONEMA HYODYSENTERIAE (SERPULINA HYODYSENTERIAE).  
 OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BRACHYSPIRA.  
 RN  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-38.  
 RC STRAIN-C5;  
 RX MEDLINE: 92307926.  
 RA KOOPMAN M.B.H., DE LEEUW O.S., VAN DER ZEIST B.A.M., KUSTERS J.G.;  
 RT "Cloning and DNA sequence analysis of a Serpulina (Treponema)  
 RT hyodysenteriae gene encoding a periplasmic flagellar sheath  
 RT protein.";  
 RL INFECT. IMMUN. 60:2920-2925(1992).  
 RN [2]  
 RP SEQUENCE OF 20-38.  
 RC STRAIN-C5;  
 RX MEDLINE: 93139764.  
 RA KOOPMAN M.B.H., BAATS E., VAN VORSTENBOSCH C.J.A.H.V.,  
 RA VAN DER ZEIST B.A.M., KUSTERS J.G.;  
 RT "The periplasmic flagella of Serpulina (Treponema) hyodysenteriae are  
 RT composed of two sheath proteins and three core proteins.";  
 RL J. GEN. MICROBIOL. 138:2697-2706(1992).  
 CC -1- FUNCTION: COMPONENT OF THE OUTER LAYER OF THE FLAGELLA.  
 CC -1- SUBUNIT: OUTER LAYER IS COMPOSED OF TWO SHEATH PROTEINS, FLAAL  
 CC (44 KD) AND FLAAL2 (35 KD) AND A CORE THAT CONTAINS THREE PROTEINS  
 CC FLAAL1 (37 KD), FLAAL2 (34 KD) AND FLAAL3 (32 KD).  
 CC  
 CC -1- SUBCELLULAR LOCATION: PERIPLASMIC FLAGELLUM.  
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X63006; G43524; -  
DR EMBL; A24079; G904393; -  
DR PIR; A43824; A43824;  
KM FLAGELLA; SIGNAL; PERIPLASMIC.  
FT SIGNAL 1 19  
FT CHAIN 20 320  
FLAAL  
FT CONFLICT 21 21 T -> S (IN REF. 2).  
FT SEQUENCE 320 AA; 35986 MW; C12634D7 CRC32;

Query Match 19.3%; Score 331.5; DB 1; Length 320;  
Best Local Similarity 28.8%; Pred. No. 3.2e-17;  
Matches 90; Conservative 63; Mismatches 129; Indels 31; Gaps 8;

QY 9 LEFLSTVLFRAQETDGLAEGSKRAEPGLVDPAELADPSSTRDLTNVYVYSGASG 68  
D 4 LEVLTSLIFLAASAYGLNST-----LIDFA-----LTGNADNLQAGEGD 43

QY 69 IVKPEDVVDGGINNWSVLTLPASRLQAYKNSVYAPAVYSESKRYAGDIIIGVYVFP 128  
D 44 TNEVYPAENLYNDNMYVWLNESARLTENRINSTYTNVDSGNGAWAGVLYVRRHP 103

QY 129 SYGSS-AMIMPKEIPYSGESGNOFL-GKGLIDNITKMEIKVSYSLGEYIDLEYLF 186  
D 104 LAANWSVALVRYELEYGAGDGTKEYGKGYIHNWGEISISSWYGRMYLISYFNL 163

QY 187 EDNMGMEYASMGTLKRGADLIMSNTYIPNISSRIIKDVPNYPLASSKMFKAERV 246  
D 164 QNEGELKSYPMGTGVYENGROVEMENREYLPNVRDVLVRE-PLYPMIPSVKLSIGF 222

QY 247 SKSHSVKNFIYVVDRLVYDKLSVSDSDISESFEKYETSGTESLRLKAHEFEK 306  
D 223 YRIQDTGGDTTYVKKDTLEYDVVVDVEEDIDDEATWOLK--TENDRK-QAIES-- 277

QY 307 RVLKLREKISIAE 319  
D 277 --ARIRQEAELRD 287

Db

RESULT 4  
ID YCXD\_BACSU STANDARD; PRT; 444 AA.  
AC 008792;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 50.8 KD PROTEIN IN SRFA4-SFP INTERGENIC REGION (ORF8).  
GN YCXD.  
OS BACILLUS SUBTILIS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
OC BACILLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168 / JH642;  
RX MEDLINE; 93360813.  
RA COSMITA P., RODRIGUEZ F., DE FERRA F., GRANDI G., PEREGO M.,  
RA VENEZA G., VAN SINDEREN D.;  
RT "Sequence and analysis of the genetic locus responsible for surfactin  
RT synthesis in *Bacillus subtilis*."  
RL MOL. MICROBIOL. 8:821-831(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-168;  
RX MEDLINE; 97124189.

RA YAMANE K., KUMANO M., KURITA K.;  
RT "The 25 degrees-36 degrees region of the *Bacillus subtilis*  
RT chromosome: determination of the sequence of a 146 kb segment and  
RT identification of 113 genes."  
RL MICROBIOLOGY 142:3047-3056(1996).  
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (POTENTIAL).  
CC -1- SIMILARITY: IN THE C-TERMINAL, TO AMINOTRANSFERASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X70356; G396486; -  
DR EMBL; D50453; G1805428; -  
DR EMBL; Z99105; E1182308; -  
DR EMBL; Z99106; E1182323; -  
DR SUBTILIST; BG10175; YCXD.  
DR PRAM; PR00392; gntP; 1;  
KM HYPOTHETICAL PROTEIN; TRANSFERASE; AMINOTRANSFERASE;  
FT BINDING 286 286 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
FT SEQUENCE 444 AA; 50755 MW; F384C1D2 CRC32;

Query Match 6.3%; Score 109; DB 1; Length 444;  
Best Local Similarity 21.6%; Pred. No. 0.68;  
Matches 75; Conservative 42; Mismatches 126; Indels 104; Gaps 15;

QY 29 SKRAPEGLVDPAELADPSSTRDLTNVVD-----YYSGASG--- 70  
D 73 SKSGQPEP--IDFATSPDP-----DVFYLDFOHCINKAIDTYKNDLFYIGPKGLPSL 125

QY 70 -----VKPEDVVDGGINNWSVLT-----TSARLQAAVKN 100  
D 126 IRLVRLKLLAQGVADNRHFIITSGVQALSLICAMPFGKKEKIAIEQGYHLMVQLE 185

QY 101 SYVAPAVYSESKRYAGDITLGVRVLPFSYSQSSAMIMPKEIPYSGESGNOFLGGLI 160  
D 186 TLGIPAIGVKTRE--GLDIKVERLFPOTESIKFFYTPRPHNP-----LGCSL 233

QY 161 DNITKMEIKVSYSLGEYIDLEYLFEDNMGMEYASMGTLKRGADLIMSNTYIPNI 220  
D 234 ERDRQ-----ELVRLAEVDVYLVEDDY-----LGDEENKKAD-----PLYAYDL 274

QY 221 SSRITKDVPNYPLASSKMFKAERVSKSHSKVKNFIYVVDRLVYDKLSVSDSDID 280  
D 275 SSHYI-----YLSFSKMFPGRLRGAVVLPALDTFTFA-----YKILN-DIDCSMI 321

QY 281 SESVFEKYETSGTESLRLKAHEFEK-RVLKLREKTI---SIAGSF 322  
D 322 SQALELLEYLKSNGMRKEREKIRDSYKERSRLHQAIRTHQOLGSGRF 368

Db

RESULT 5  
ID Y686\_METUA STANDARD; PRT; 580 AA.  
AC 058099;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL PROTEIN MJ0686.  
GN MJ0686.  
OS METHANOCOCCLUS JANNAASHII.  
OC ARCHAEA; EUDYARCHAEOTA; METHANOCOCALLES; METHANOCOCALACEAE;  
OC METHANOCOCCLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE; 96337999.

RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
 RA KERAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,  
 RA OVERBEK R., KIRKESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,  
 RA SCOTT J.L., GOGGAGEN N.S.M., WEIDMAN J.F., FUHRMAN J.L., NGUYEN D.,  
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,  
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,  
 RA KLEINK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.,  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RT SCIENCE 273:1058-1073(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U67515; G149499; -  
 DR TIGR: M0686; -  
 KM HYPOTHEICAL PROTEIN  
 SQ SEQUENCE 580 AA; 68392 MW; 5C3F000E CRC32;  
 Query Match 6.3%; Score 108; DB 1; Length 580;  
 Best Local Similarity 21.7%; Pred. No. 1.1; Mismatches 66; Indels 98; Gaps 11;  
 Matches 56; Conservative 38;  
 QY 112 SKRYAGDTLLGVRLVFPYSQSSAMIPPKIPYSGESGNOFLGKLDINIKTKIEKV 171  
 DB 376 SERYSGST-----ESKMKDKFEKVSQWENNKI--PEVENIKI--EIQY 417  
 QY 172 SVYSLGEYDLEVPEDMNGMEYAYSMGLKFKGADLWSPN--YIPNISRIRKND 228  
 DB 418 GAUUL-----KKVSSGSSKYGVEFLFKPKPMDYKRP-----DN 453  
 QY 229 VENVYPLASSKMKFKAF-----RVSKSHSKYKNFIYK 262  
 DB 454 IAYKKEEDHHPKGLRKNKGISNEYIDSVLNTPIIDETNKKIKSKSKS-----YK 507  
 QY 263 DLRLVLDKLSVDSIDSDID-----SESVFYKETSGETSLKRLAHEFKRVLK 311  
 DB 508 EMEIDKNGKLGSEDAVNVKVELKGLFNEEMFELIRMT--DLSLKDIEENFNFIEL 566  
 QY 312 RKSTIAEGSFQNFVEKI 329  
 DB 567 REKL-----ILEKI 575  
 RESULT 6  
 ATPL\_METJA STANDARD; PRT; 695 AA.  
 ID ATPL\_METJA  
 AC 057675;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE PROBABE ATP SYNTHASE SUBUNIT I (EC 3.6.1.34).  
 GN MJ0222.  
 OS METHANOCOCCUS JANNASCHII.  
 OC ARCHAEA; EUBRYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;  
 CC METHANOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE: 96337999.  
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,  
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,  
 RA KERAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,  
 RA OVERBEK R., KIRKESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,  
 RA SCOTT J.L., GOGGAGEN N.S.M., WEIDMAN J.F., FUHRMAN J.L., NGUYEN D.,  
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,

RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,  
 RA KLEINK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.,  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RT SCIENCE 273:1058-1073(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U67478; G1590960; -  
 DR TIGR: MJ0222; -  
 KM HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE.  
 FT TRANSMEM 392 412 POTENTIAL.  
 FT TRANSMEM 425 445 POTENTIAL.  
 FT TRANSMEM 497 517 POTENTIAL.  
 FT TRANSMEM 534 554 POTENTIAL.  
 FT TRANSMEM 556 576 POTENTIAL.  
 FT TRANSMEM 605 625 POTENTIAL.  
 FT TRANSMEM 627 647 POTENTIAL.  
 SQ SEQUENCE 695 AA; 76953 MW; 803B3BA3 CRC32;  
 Query Match 6.2%; Score 106.5; DB 1; Length 695;  
 Best Local Similarity 23.5%; Pred. No. 1.9; Mismatches 108; Indels 125; Gaps 19;  
 Matches 84; Conservative 41;  
 QY 68 GIVPEPMVVDLGINNVSVLLTPSARLQAYKNSVAVAVKSSKRRAGDTLLGVRL 127  
 DB 32 GIVELCDLSEKLEDELMKTLSPSSAD-YVKN-VISLTK-----AG-----RILD 76  
 QY 127 -FPYSQSSAMT-----MPPFKIPYSGESGNOFLGKLDINIKTKIEKV 161  
 DB 77 MESSVSQKETSIKDLINKPKYBEKKVSNFSQYVIDAYEKUNELSKREVDPARLSELD 136  
 QY 162 NIKT-MKEIKVSV-YSLGEYDLEVPEDMNGMEYAY-----SKGLKFKGADLI 210  
 DB 137 NKKSKLQLEKQISYTLKGLFEPDLKYL-----SGGEVVFAGASVPEKIKGEIK-----AELD 188  
 QY 211 WSNPYINIS-SRIIKDDVPNYPPLASSKMKFKAFVSKSHSKYKNFIYKDLRLVLD 269  
 DB 189 KVADGYIGIFSGSEFEKDKIRPIVYTLKEKLENV-----LSEIRKEEFERYDI----- 240  
 QY 270 KLSVDSIDSDIDS-----ESVFKYETSGTESLRLK--AHETFRVLDLREKISI- 318  
 DB 240 -----SDVEGTPSEALSKIESLKAIESERNLSIEKTLAKQWKEKELLAYVELLSIE 292  
 QY 318 -----ABGSFQNF-----FVEKTESKEPESSP 339  
 DB 293 KARGDAVSQFGKTDRTYIEMVVPARDAEKAKSLIENSADGFAFVEITPEDEPEKIP 350  
 RESULT 7  
 LITB\_BACSU STANDARD; PRT; 705 AA.  
 ID LITB\_BACSU  
 AC 002113;  
 DT 01-JUL-1993 (REL. 26, CREATED)  
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)  
 DE AMIDASE ENHANCER PRECURSOR (MODIFIER PROTEIN OF MAJOR AUTOLYSIN).  
 GN LITB OR CWBA.  
 OS BACILLUS SUBTILIS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
 CC BACILLUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;



Db 374 NKAOSLPFNSRASIL-----IMLIDGP-----TEGYTDRSOLIKNVRDAIRGRFPLY 423  
Qy 175 SGEYDELEVLPEDMNGME-----YAYSMGLTKFKGMADLJNSPNYIPNISRIK 226  
Db 424 NIGFGHGVEMNLEVALALNNGRAQRIYEDHDSAOLOOFGYOV-ANP-----LIK 473  
Qy 227 DDVNPY-----LASSKMRKAFR-----VKSSSHKKVKNFIFYKDLRVLYDKLSVI 275  
Db 474 DVELQPADAVLALTOHRRKQYEGSEITVAGRIADNKLSE-----KADVQA 521  
Qy 276 DSD-----IDSEVFEVYETSG-----TESLRKLAHETFRVYKLRKISIAE 319  
Db 522 SGDQGFYTTCLVDEEMKRLLOERGHMLENYERLWAYLITLOELLAKRMLKMAE 576

## RESULT 9

XYLB\_BACSU STANDARD; PRT: 499 AA.  
ID XYLB\_BACSU  
AC P39211;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE XYLUOSE KINASE (EC 2.7.1.17) (XYLUOKINASE).  
GN XYLB.  
OS BACILLUS SUBTILIS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;  
OC BACILLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BORHERT S., KLEIN C., PIKSA B., HAMMELMANN M., ENTIAN K.D.;  
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE OF 1-8 FROM N.A.  
RX MEDLINE: 85297769.  
RA WILHELM M., HOLLENBERG C.P.;  
RT "Nucleotide sequence of the Bacillus subtilis xylose isomerase gene:  
RT extensive homology between the Bacillus and Escherichia coli  
RT enzyme.";  
RL NUCLEIC ACIDS RES. 13:5717-5722(1985).  
CC -1- CATALYTIC ACTIVITY: ATP + D-XYLUOSE -> ADP + D-XYLUOSE  
CC 5-PHOSPHATE.  
CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCOKINASE /  
CC GLYCEROKINASE / XYLUOKINASE FAMILY.  
CC CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC CC  
DR EMBL: U66480; G1750125; -;  
DR EMBL: X02795; G580946; -;  
DR EMBL: Z99113; E1183420; -;  
DR SUBTILIST; BG10807; XYLB.  
DR PROSITE; PS00445; FGXY\_KINASES\_2; 1.  
DR PROSITE; PS00933; FGXY\_KINASES\_1; 1.  
DR PFAM; PF00370; FGXY; 1.  
KM TRANSFERASE; KINASE; XYLOSE METABOLISM.  
SQ SEQUENCE 499 AA; 55417 MW; 1E3E31A1 CRC32;

Query Match 6.1%; Score 104; DB 1; Length 499;  
Best Local Similarity 20.9%; Pred. No. 1.8; Indels 66; Gaps 14;

Matches 68; Conservative 57; Mismatches 135; Indels 66; Gaps 14;

Qy 30 KRAEPGELVDFAEALRDSSTRDLTINYDYYVYSGASG-----YKPE---DMVVDLGI 81  
Db 141 KEHEP-ELFKKTAVALPLPDYVRFRTGVITHESDACTLLHTRKMSNDICQIGI 199  
Qy 82 NMSVLLTPSARLQAVYKNSVVAFAVYKSESKRYAGDTILGVAVLFPYSQSASAMIMPF 141

Db 200 -----SADICPELVESH-----CVSGLLPVHAKTGLLE--- 230  
Qy 142 KIPFSGESGNO--ELAGLIDNIKTMEIKVSYSLGEYIDLE-----VLFEDMNGME 193  
Db 230 KTKYAGADNACGAIAGIISGKTLCSITGSVILSYEEKEREDFGKHFNNHKKD 289  
Qy 194 YAYSMGLTKFKGMADLJNSPNYIPNISRIKDDVNPYPLASSKMRKAFRVSK--SHS 251  
Db 290 SEYTMGVTLAAGYS-LDFEKRTRFAPNESFEQLQGVFAIPGANGLLYTPYLVGERPFA 348  
Qy 252 -SKVNFIFYKDD-----LRVLYDKLSVSDSID-----SESVFVYETSG---TES 295  
Db 349 DSSIGSLIGMDGAAHNRHFRALMEGITFSLHESIELFREAGRSVHTVSIGGAKNDT 408  
Qy 296 LRKLAHETFRVYKLRKISIAEGS 321  
Db 409 WLQMDLFPNTRVITLNEQGPAMGA 434

## RESULT 10

HS71\_LYCES STANDARD; PRT: 650 AA.  
ID HS71\_LYCES  
AC P24629;  
DT 01-MAR-1992 (REL. 21, CREATED)  
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HEAT SHOCK COGNATE 70 KD PROTEIN 1.  
GN HSC-1.  
OS LYCOPERSICON ESCULENTUM (TOMATO).  
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMERIOPHYTA; TRACHEOPHYTA;  
OC EUPHYLLIOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;  
OC ASTERIDAE; SOLANACEAE; SOLANACEAE; SOLANACEAE; SOLANUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-CV, MTL VF36; TISSUE--PISTIL;  
RX MEDLINE: 91370894.  
RA LIN T.Y., DUCK N.B., WINTER J., FOLK W.R.;  
RT "Sequences of two hsc 70 cDNAs from Lycopersicon esculentum.";  
RL PLANT MOL. BIOL. 16:475-478(1991).  
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
CC CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC CC  
DR EMBL: X54029; G19256; -;  
DR PIR: S14949; S14949.  
DR PROSITE; PS00297; HSP70\_1; 1.  
DR PROSITE; PS00329; HSP70\_2; 1.  
DR PROSITE; PS01036; HSP70\_3; 1.  
DR PFAM; PF00012; HSP70; 1.  
DR HSP; P19120; INGI.  
KM ATP-BINDING; HEAT SHOCK; MULTIGENE FAMILY.  
SQ SEQUENCE 650 AA; 71287 MW; 98C96C58 CRC32;

Query Match 6.0%; Score 103; DB 1; Length 650;  
Best Local Similarity 18.8%; Pred. No. 3.1;

Matches 68; Conservative 57; Mismatches 124; Indels 112; Gaps 13;

Qy 43 ELADPSSTRDLTINYDYYVYSGASGIYKPEDMVVDLGINMSVLLTPSARLQAVYKNSV 102  
Db 61 QVALNPINTVDFAKRLIGRRSDAS---VQEDMKL-----WPKVIFGPQDK----- 105  
Qy 103 VAPAVY-----KSESKRYAGDTILGVAVLFPYSQSASAMIMPFK----- 143  
Db 105 -PMIVTVYTKGEKEFEFAEEI-----SSWVLTIMKEIAEALFGLSTYKNAVYT 149  
Qy 143 IPFYSGESGNOFL-GKGLIDNIKTMEI-KVSYSLGEYIDLE-----VLFEDMNGM 192



DB 150 VPAYFNDOSQATKAGVIGSLNVMRLINEPTAAALAYGLDKRATSGAKENVLIFDGGG 209  
QY 193 EYAYSMGTLK-----FKWADLIWSPNYIPNISRITIDVPNTPLASSK 239  
DB 210 TFDVSLITIEGFEYKATAGDTHLGGEDPDNRKVNHFVEFRKAKKDDITGNPRALRL 269  
QY 240 RFAFRVSKSHSSKVNKFI-----FYVDRLVLYDKLSVSI-----D 276  
DB 270 RTACERAKRRTLSSTAQTIEIDSLVEGVDFYSTTRAFEFELMMNDLFRKCMPEYKCLRD 329  
QY 277 SDDSDSEVFYETSGTESRLKAHETFRKVLKREKISIAEGSFONYEKEISEKPEE 336  
DB 330 AKMDKSTVHDVVLVGSGSTRIPKVO-----QVAMTFNFKGKELCKSIMPDE 374  
QY 337 S 337  
DB 375 A 375

RESULT 11  
ITHL.MESAU  
ID ITHL.MESAU STANDARD; PRT; 914 AA.  
AC P97278;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE INTER-ALPHA-TRYP SIN INHIBITOR HEAVY CHAIN H1 PRECURSOR (ITI HEAVY CHAIN H1).  
GN ITIH.  
OS MESOCRICEUS AURATUS (GOLDEN HAMSTER).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; CRICETINAE; MESOCRICEUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RA NAKAIZANI T., SUZUKI Y., YAMAMOTO T., SINOHARA H.;  
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
-1- SUBUNIT: INTER-ALPHA-TRYP SIN INHIBITOR CONSIST OF A LIGHT CHAIN AND AN HEAVY CHAIN. THERE ARE THREE DIFFERENT HEAVY CHAINS.  
-1- SIMILARITY: BELONGS TO THE ITIH FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; D89285; D1014635; .  
DR PRAM; PFO0092; YWA; 1  
KM SERINE PROTEASE INHIBITOR; REPEAT; SIGNAL; MULTIGENE FAMILY;  
KW GLYCOPROTEIN.  
FT SIGNAL 1 914  
FT CHAIN ? 914  
FT FT ? 914  
FT CARBOHYD 288 288  
FT CARBOHYD 291 291  
FT CARBOHYD 591 591  
FT CARBOHYD 753 753  
SO SEQUENCE 914 AA; 101785 MW; B7406361 CRC32;

Query Match 6.0%; Score 102.5; DB 1; Length 914;  
Best Local Similarity 19.1%; Pred. No. 5.3;  
Matches 66; Conservative 59; Mismatches 119; Indels 101; Gaps 15;

QY 16 VLFAOETDGAEGSKRAPELVLDFAELARDPSTRDLJNVDYYSAGSIVREED- 75  
DB 296 LVFVYIDISGSEKOGKAVQTEALKIL-----GVKRGDS 330  
QY 75 ---MVVDLGINNSVLLTPS--ARLQA---YVKNVAVAPAVKSESKRYAGDTILGVRL 126

DB 331 FDLVLRGSRVQSKGSLVPATQANLQAADFVFRFSLAGT-----NLNGILRGIEIL 384  
QY 127 -----FPSYSSANIMPFKIPFYSGESGNQFLGKGLINDIKTKRKIVSV-----Y 174  
DB 385 NKAQSHPELSSPASTL-----TMLTDEP-----TEGETDRSQILKNVNAALRGPPLY 434  
QY 175 SLAGEIDLEVLFDPMNGMEYASMGTLKFKGNADLIWSN-----PNYIPNISRILKD 227  
DB 435 NLGFGDLDFNLELVSMENS-----GNMRRIYEDDADQOLGFIYNOVANPLLD 485  
QY 228 DVNYP-----LASSKMFRAF-----RVSKSHSSKVNKFIYVDRLVLYKLSV 273  
DB 486 VELQYPODSVLSLTOHRHKQYDGESEIYVAGRIADKLKSTFK-----ADVRRGEROEP 539  
QY 274 SDDSDSEVFYETSG-----TESRLKAHETFRKVLKREKI 315  
DB 540 KATCLVDEEMKRLRLRERGHMLNHEVRLWALYTLIOELLAKRKRM 584

RESULT 12  
HS7C\_TRYB  
ID HS7C\_TRYB STANDARD; PRT; 676 AA.  
AC P20030;  
DT 01-FEB-1991 (REL. 17, CREATED)  
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)  
DE HEAT SHOCK COGNATE HSP70 PROTEIN.  
OS TRYPANOSOMA BRUCEI BRUCEI.  
OC EUKARYOTA; EULENZOZA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90377290.  
RA LEE M.G.-S., POLYERE R.I., VAN DER PLOEG L.H.T.;  
RT "Evidence for segmental gene conversion between a cognate hsp 70 gene and the temperature-sensitive transcribed hsp70 genes of Trypanosoma brucei."  
RL MOL. BIOCHEM. PARASITOL. 41:213-220(1990).  
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; M32139; G162125; .  
DR PIR; A45515; A45515.  
DR PROSITE; PS00297; HSP70\_1; 1.  
DR PROSITE; PS00329; HSP70\_2; 1.  
DR PROSITE; PS01036; HSP70\_3; 1.  
DR PFAM; PFO0012; HSP70; 1.  
DR HSSP; P19120; INCG.  
KW ATP-BINDING; HEAT SHOCK; MULTIGENE FAMILY.  
SO SEQUENCE 676 AA; 73694 MW; 489BCDAF CRC32;

Query Match 5.9%; Score 101.5; DB 1; Length 676;  
Best Local Similarity 23.2%; Pred. No. 4.2;  
Matches 73; Conservative 44; Mismatches 112; Indels 85; Gaps 16;

QY 63 YSGASGIVKPEDMVVDLG-----INNSVLLTPSARLQAYVKNVYAP 105  
DB 3 YEGAIGI-----DLGTYSCVGVQNERVEIIANDQGNRTTPS--YVAFVNNELVVG 52  
QY 106 AVYKSESKRYAGDTILGV--RVLFPSYSSQ--SAMIIMPFKIPFYSGESG-----OFLG 156  
DB 53 DAAKSHAAKSNQVYIDAKKLIGRKFSDSVYQSDMKHWPKEV--DEKKGAVMRVHLG 110  
QY 157 KGLIDNITKTKRKIVSVYSGLYEIDLEVLFDPMNGMEYASMGTLKFKGNADLIWSNPY 216







GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 17:16:47 ; Search time 543.6 Seconds  
(without alignments)  
122.860 Million cell updates/sec

Title: US-09-004-395-3

Perfect score: 21  
Sequence: 1 atgaaaggaaggaagctaaagt 21

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl: \*  
1: gb\_dal: \*  
2: gb\_daz: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pil: \*  
8: gb\_pil2: \*  
9: gb\_prl: \*  
10: gb\_prl2: \*  
11: gb\_prl3: \*  
12: gb\_prl4: \*  
13: gb\_prl5: \*  
14: gb\_prl6: \*  
15: gb\_prl7: \*  
16: gb\_prl8: \*  
17: gb\_prl9: \*  
18: gb\_prl10: \*  
19: gb\_prl11: \*  
20: gb\_prl12: \*  
21: gb\_prl13: \*  
22: gb\_prl14: \*  
23: gb\_prl15: \*  
24: gb\_prl16: \*  
25: gb\_prl17: \*  
26: gb\_prl18: \*  
27: gb\_prl19: \*  
28: gb\_prl20: \*  
29: gb\_prl21: \*  
30: gb\_prl22: \*  
31: gb\_prl23: \*  
32: gb\_prl24: \*  
33: gb\_prl25: \*  
34: gb\_prl26: \*  
35: gb\_prl27: \*  
36: gb\_prl28: \*  
37: gb\_prl29: \*  
38: gb\_prl30: \*  
39: gb\_prl31: \*  
40: gb\_prl32: \*  
41: gb\_prl33: \*  
42: gb\_prl34: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	13271	2 AE001168	AE001168 Borrelia

2	21	100.0	1555	2	BRU62900	U62900 Borrelia bu
3	20	95.2	3885	7	LEPRR	X73156 L.esculentu
4	18.4	87.6	33045	36	CET25C12	Z65566 Caenorhabdi
5	18	85.7	112621	11	AC004891	AC004891 Homo sapi
6	18	85.7	73360	11	HSU66083	U66083 Human cont1
7	17.8	84.8	2794	4	XLNU1974	U19974 Xenopus lae
8	17.8	84.8	40397	7	SC9959	Z49939 S.cerevisia
9	17.8	84.8	7912	10	HS7553D10	AL094651 Human DNA
10	17.8	84.8	15361	34	HS11141	AL031656 Homo sapi
11	17.8	84.8	111641	34	HS232N11	AL031656 Homo sapi
12	17.4	82.9	99856	7	ATF24C24	AL031656 Homo sapi
13	17.4	82.9	12389	7	SC9959	Z49939 S.cerevisia
14	17.4	82.9	14989	7	SC9959	Z49939 S.cerevisia
15	17.4	82.9	63922	8	T9A4	AF096373 Arabidops
16	17.4	82.9	126263	9	HS127B14	Z93928 Human DNA s
17	17.4	82.9	299719	34	AC006780	AC006780 Caenorhab
18	17.4	82.9	299719	34	AC006858	AC006858 Caenorhab
19	17.4	82.9	294136	34	AC006901	AC006901 Caenorhab
20	17.4	82.9	96863	35	AC007259	AC007259 Arabidops
21	17.4	82.9	23536	36	CET25C12	U97013 Caenorhabdi
22	17.4	82.9	28374	36	DMAC001666	AC001666 Drosophila
23	17.4	82.9	85260	37	AC001661	AC001661 Drosophila
24	17.4	82.9	75609	42	AC006456	AC006456 Homo sapi
25	17.4	81.0	216293	11	HTCRBCR9	AF029308 Homo sapi
26	17.4	81.0	7102	17	TACTS8G	J04340 Tacaribe vi
27	17.4	81.0	189742	35	AC007388	AC007388 Homo sapi
28	16.8	80.0	13205	2	U67516	U67516 Methanococ
29	16.8	80.0	976	7	LEPARGENE	Y15490 Lycopersico
30	16.8	80.0	45731	7	SPBC3H7	AL031261 S.pombe c
31	16.8	80.0	133840	8	AC003981	AC003981 Complete
32	16.8	80.0	1235	8	AF059489	AF059489 Lycopersi
33	16.8	80.0	88543	8	ATAC004680	AC004680 Arabidops
34	16.8	80.0	84592	8	ATAC007584	AC007584 Arabidops
35	16.8	80.0	89469	8	ATF18B3	AL049862 Arabidops
36	16.8	80.0	1500	8	CAU95611	U95611 Candida alb
37	16.8	80.0	483	9	D45028S12	D45029 Human DNA f
38	16.8	80.0	84464	10	HS782D21	AL022399 Human DNA
39	16.8	80.0	2198	10	HS133525	X69814 H.sapiens D
40	16.8	80.0	134649	11	AC004047	AC004047 Homo sapi
41	16.8	80.0	163712	11	AC004065	AC004065 Homo sapi
42	16.8	80.0	112456	11	AC004614	AC004614 Homo sapi
43	16.8	80.0	186044	11	AC004831	AC004831 Homo sapi
44	16.8	80.0	143834	11	AC004960	AC004960 Homo sapi
45	16.8	80.0	121803	42	AC004835	AC004835 Homo sapi

## ALIGNMENTS

RESULT 1  
AE001168  
LOCUS AE001168 13271 bp DNA BCT 15-DEC-1997  
DEFINITION Borrelia burgdorferi (section 54 of 70) of the complete genome.  
ACCESSION AE001168 AE000783  
NID 92688598  
VERSION AE001168.1 GI:2688598  
KEYWORDS  
SOURCE  
ORGANISM  
Lyme disease spirochete.  
Borrelia burgdorferi  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
burgdorferi group.  
1 (bases 1 to 13271)  
REFERENCE  
AUTHORS  
Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A.,  
Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K.,  
Gwinn, M., Dougherty, B., Tomb, J.-F., Fleischmann, R.D.,  
Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J.,  
Salzberg, S., Hanson, M., Van Vugt, R., Palmer, N., Adams, M.D.,  
Gocayne, J.D., Weidman, J., Uterback, T., Watthey, L., McDonald, L.,  
Artach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K.,  
Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.  
Genomic sequence of a Lyme disease spirochete, Borrelia  
burgdorferi  
JOURNAL Nature 390 (6660), 580-586 (1997)

FEATURES	source
TITLE	2 (bases 1 to 13271)
JOURNAL	Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Ladhira, R., White, O., Ketchum, K.A., Dodson, R., Hickley, E.K., Gwin, M., Dougherty, B., Tomb, J.-F., Fleischmann, R.D., Richardson, D., Peterson, J., Kertész, A., Quackenbush, J., Salzberg, S., Hanson, M., Van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J.D., Weidman, J., Uitterlinden, T., Matthey, L., McDonald, L., Artach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.
DESCRIPTION	Submitted (12-DEC-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
LOCATION	Location/Qualifiers
1..13271	/organism="Borrelia burgdorferi"
/db.xref="taxon:139"	
66..1100	
/gene="BB0668"	
66..1100	
/gene="BB0668"	
/note="similar to GP:1575447 percent identity: 98.83; identified by sequence similarity; putative"	
/codon_start=1	
/transl_table=1	
/product="Flagellar filament outer layer protein (flaA)"	
/protein_id="AAC67025.1"	
/db.xref="PID:92688608"	
/db.xref="GI:2688608"	
/translation="MVMKRRKAKILFFILSTVFAETOGSLAGSKRAEGLVLPD AELARDPSRDLTNYVDYYSASGIVREEDVYDGLNNVSLTPSRKLAQYKR NSVYAPVAKSEKRIAGDITLGVLPSTSSQSAMTPKPIETFGSGNOFLK GLDINIKTMEIKVSVSLGEIDLEVLFDENOMEXAYSGILTFKGMADLMSNPB YIPISRSIKIDVNPYPLASCKMRFAFRVSKSHSKDEQNFIFVADLRVYLKLSV SIDSDISESYFKVETSGFESLRLKAHEFFKRVKLKREKISMPESGQFVERIES ESPESSPKN"	
1168..3762	
/gene="BB0669"	
1168..3762	
/gene="BB0669"	
/note="similar to PID:1113815 SP:044737 GB:AE000783 percent identity: 99.31; identified by sequence similarity; putative"	
/codon_start=1	
/transl_table=1	
/product="Chemotaxis histidine kinase (cheA-2)"	
/protein_id="AAC67024.1"	
/db.xref="PID:92688607"	
/db.xref="GI:2688607"	
/translation="MELIDLENEELLVFPEEAQNVLDLEENMSLEDDPNNSDTID EIPFAARLKGSSASLMMELSDPTHVYEDFEDAIRGKAVINNDVYDLSLDYIKR EMALARDKYVLDNDLSKRLKQFLVYDQITIKRFGDSINNCLSDSELEIR EGDLIGVLRISVFNSSNSVEVNSGKLFENLKNLGSVLTPIPEQDITIEDKFLK RVDTLYGLSDIEGVKSLDSLNTLESYLVDENPKMELKLADEIKDVIDDSVNLN DNPFETDEISDLLEVENQKLFVRLDVFQNDPMATISGLMQLAKSLGKIRSPIS DSESLDAKDFEDFYIYLLISMTSESIARKNLDPVASHPEIKVNSLESKSVLKD EDAAPEKNNKIKKSPISVNLIRDSKIDVILDLVEASHINSLSFYAKNOISMTLEPY NENYVQESFOENPLDLYKFKDAGITIEDLESEHNSLSFYAKNOISMTLEPY NSPFRLLQNKRTISGRLSRIITDLESVLTAKRLMLPISIRSFIRVVRDSKRLKTV LNKKEGETELDKSVIDDVLPALMHCVRNSMDHLEVEEVKRGSKAGTILIRANK EGNVLSIEIDDGIGIDPKVIRKLIRKIGKTIKEDAIYSDPELMLIFAPGSHVQVYT DLSGCVGLDVAKSIEKINGTILVESEIGIGTIFKTLPLTVIIGGLVKSGETYV VILPNVLEHRIETENDIKLENYHEVYNLADEVYSVLRDLKLNIRDSLSLEKPII VVNSNMKIALIVDSILIGEDFVVKPIKDKFSSAGIVGATTLGNKAVYLLIDYKLF DLDQTE"	
3763..5169	
/gene="BB0670"	
3763..5169	
/gene="BB0670"	
/note="similar to GP:1881576 percent identity: 98.93; identified by sequence similarity; putative"	
/codon_start=1	
/transl_table=1	

gene  
CDS

/product="purine-binding chemotaxis protein (chew-3)"  
/protein\_id="AAC67023.1"  
/db\_xref="pid:g2688606"  
/db\_xref="gi:2688606"  
/translation="MQAIKEYFGPKIIDDKNSKSLTNFDEKVVSFELGSDHYLDVIM  
QYKEIRKSNSFYVPNAKKYVAGLDNLNGEIIPIIDILINMNFENKDIEDIVLKNR  
DOLLGVLTIDNNVYSIDSSLIDPHEVLSDSLINYIKGVYNELLYLLWFKRIK  
VYGEERELKRGONFEKRSDPVSKDMLDLIENSNDSESKVSRKNDGNENSTLNKG  
TAAPLENLRKRLKYAFSPNATLVNDVLEKRYGVFNMYDDYLPCDSFLNFPYRKSSG  
NIMGDCIEERNELVEBSRLFPMANNLSIFNVLEIGCGSGETALANALSETIYKFRP  
KLTIADNDISKAVETSRLVFSESEIGSETISRNFESDPGVYFKSELISNLVPEYS  
DALFDLDNLGMVELKDYLCFLDSKDQIILLINTASKAIGKALLVLDNEELNNVDY  
FIKERSTKYFNRKY"

gene  
CDS

5181..5666  
/gene="BB0671"  
5181..5666  
/note="similar to GP:1408273 percent identity: 98.76;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis operon protein (chev)"  
/protein\_id="AAC67022.1"  
/db\_xref="PID:g2688605"  
/db\_xref="GI:2688605"  
/translation="MRIDYEPLFDAASSVLDMLVENIEMGRPGKSIINOKIGVSV  
VYVGAGSVEGSIIIDMDILETALFVASKLFEEDPDDEETREMAVALTEVENIIIA  
GNVFYTLAKGVDFITPPARFYGENMKISNKSGDALIVPSLPDGKIIEVNIAIRBR  
V"

gene  
CDS

5706..6146  
/gene="BB0672"  
5706..6146  
/gene="BB0672"  
/note="similar to GP:1408274 percent identity: 97.95;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis response regulator (chev-3)"  
/protein\_id="AAC67021.1"  
/db\_xref="PID:g2688604"  
/db\_xref="GI:2688604"  
/translation="MIQKTIAADSSSKPRGINDTGIPENVLIIVDSVTFVKOIQH  
ETSEGFNIIDDAAGEAVIKRYKHYPIDIIVTDIMPKKDGTCSLSIMEPFKNARNR  
VIMISALKEQLVDRCLTKGAKFTIVPLDKAVLQRMVSFVR"  
complement(6187..6702)  
/gene="BB0673"  
complement(6187..6702)  
/note="similar to GP:1408275 percent identity: 97.56;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC67020.1"  
/db\_xref="PID:g2688603"  
/db\_xref="GI:2688603"  
/translation="MKIOIITMLALDLPFNARULDISIKRADEIKKYSYNLLI  
EKETYNPTISEIKNIKTLTEHFVKSTLMKNTNLSINSYKEANKYLIQSESLDKR  
FLKYFIKRNLIKNIFFKSHSLITYRKGYKLEYLIENNAEPKILFINITYFLNKDKR  
ISNMEIFEPR"

gene  
CDS

complement(6699..7745)  
/gene="BB0674"  
complement(6699..7745)  
/gene="BB0674"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0674"  
/protein\_id="AAC67027.1"  
/db\_xref="PID:g2688610"  
/db\_xref="GI:2688610"  
/translation="VWKFSIFLKALIIIFSIFELLIELSLIIIFLPYKIRPALIFGEF

LEPTIFIFLYKITKAYLSORLEIYVNNLFDIICHLIPLAFYSSYQKNTIVAE  
TINIPMLSLFKRLRLRNDLIEIYNSKKNLILAFATFMSLLIPETFI  
IISSEKIVNISPEKOEENIINISINERKAYIKKYPFILIIEKODIISKDEIV  
YSPSEKRVIEKTEKTYIDYLOKSDSLIGLFTLFPSTFILNFKPKFASL  
NPILMKKILIODPLEYKRIQIPTLSEKRYELAKSNLLKKNLSKRSKPILEI  
EKVKIINKMOEIK"  
/complement (7739, .8602)  
/gene="B80675"  
/complement (7739, .8602)  
/gene="B80675"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region B80675"  
/protein\_id="AAC67026.1"  
/db\_xref="PID:g2888609"  
/db\_xref="GI:2688609"  
/translation="MNTKPYEPLYHYLFNHSIKSLAEKEIEILNLYKKNKTTA  
TFIKNDESEIKDIQYKDKTDIMIPVLGSLAIDFNIKPLFSEKELKNDLNI  
FNFVKVNSLRKEEFYNTISNGYIFNKLPEGKNSYIYLIOKKNALYSSDI  
KNYKILLIKLVIVIKYCFKGIETLTNKNIESKATSNDDPDEKTAIIIESFK  
YETIQTSPISTLAIATFARARPTKYNVFKGTIGIDESWFSIKSGSKRDIIRIK  
ELSEIAKVNKM"  
/complement (8599, .9261)  
/gene="B80676"  
/complement (8599, .9261)  
/gene="B80676"  
/note="similar to GB:LA2023 SP:P44755 PID:1004013  
PID:122502 PID:1204815 percent identity: 25.58;  
identified by sequence similarity; putative"  
/codon\_start=1

Query Match 100.0%; Score 21; DB 2; Length 13271;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atgaaaggaagctaaagt 21  
|||||  
Db 75 ATGAAAGGAAGCTAAAGT 95

RESULT 2  
BB062900 1655 bp DNA BCF 15-JAN-1997  
LOCUS BB062900  
DEFINITION Borrelia burgdorferi flagellar filament outsheath protein (flaA)  
gene, complete cds, and chemotaxis histidine kinase (cheA) gene,  
partial cds.  
ACCESSION U62900  
NID 91575445  
VERSION U62900.1 GI:1575445  
KEYWORDS Lyme disease spirochete.  
SOURCE Borrelia burgdorferi  
ORGANISM Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
burgdorferi group.  
REFERENCE 1 (bases 1 to 1655)  
AUTHORS Ge.Y. and Charon,N.W.  
TITLE An unexpected flaA homolog is present and expressed in Borrelia  
burgdorferi  
JOURNAL J. Bacteriol. 179 (2), 552-556 (1997)  
MEDLINE 97144545  
REFERENCE 2 (bases 1 to 1655)  
AUTHORS Ge.Y.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-1996) Yigong Ge, West Virginia University,  
Microbiology, HSCN, Morgantown, WV 26506, USA  
FEATURES  
source  
1. 1655  
/organism="Borrelia burgdorferi"  
/strain="212"  
/db\_xref="taxon:139"  
/clone="pwl and pw2"

CDS  
<1..395  
/function="unknown"  
/note="orfa"  
/codon\_start=3  
/transl\_table=11  
/protein\_id="AAC44769.1"  
/db\_xref="PID:g1575446"  
/db\_xref="GI:1575446"  
/translation="DNLFKRVYFPHSSKGFVANLRYLRDEONLKONLDLVKDFLLG  
SNRGSFQPLSDSRFLYSFLKNGVYVNSRPFYDSFNNGDI NESFDKYNVFNANSL  
IKTRNRYEKIKIILIVGCLIKES"  
473..1498  
/gene="flaA"  
473..1498  
/gene="flaA"  
/function="putative flagellar filament outsheath protein"  
/codon\_start=1  
/transl\_table=11  
/product="flaA protein"  
/protein\_id="AAC44770.1"  
/db\_xref="PID:g1575447"  
/db\_xref="GI:1575447"  
/translation="MKRKASILFFLLSTVLFNQETDGLAEGSKRAEPGLVDPMEL  
ARDSSSTRDLINVDYVYSGAGIVKPEDMVVDLGINMSVLLTPSARLQAVKNSV  
VAPAVVSESKRYAGDTILGVRVLPFSYSSSSAMIMPPIFYSGESGQFQGLI  
DNIKTKMEIKVSYSLGVEIDLVLFEDMNGMAYSMGTLKFGNADLIWSPNPI  
NISRIIKDDVPNYPVLAASSKMRKARFVSKSHSKNFIYVYKDLRVYLDKISYID  
SDIDSSEYFVYVETSGTESLRKLAHETFRVYKLAKKISIAEGSFQNFPEKIESEK  
EESFKN"  
1566..1655  
/gene="cheA"  
1566..>1655  
/gene="cheA"  
/function="chemotaxis histidine kinase"  
/codon\_start=1  
/transl\_table=11  
/product="CheA protein"  
/protein\_id="AAC44771.1"  
/db\_xref="PID:g1575448"  
/db\_xref="GI:1575448"  
/translation="MELIDLENELLGVFEERQNLVDILENI"

BASE COUNT 560 a 182 c 308 g 605 t  
ORIGIN

Query Match 100.0%; Score 21; DB 2; Length 1655;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atgaaaggaagctaaagt 21  
|||||  
Db 473 ATGAAAGGAAGCTAAAGT 493

RESULT 3  
LEPUR 3885 bp DNA PLN 03-NOV-1994  
LOCUS LEPUR  
DEFINITION L.esculentum polyubiquitin repeats.  
ACCESSION X73156  
NID 9312159  
VERSION X73156.1 GI:312159  
KEYWORDS direct repeat; inverted repeat; polyubiquitin gene; ubiquitin.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Asterales; Solanales; Solanales; Solanales; Solanaceae; Solanum; Potatoc;  
Lycopersicon.  
REFERENCE 1 (bases 1 to 3885)  
AUTHORS Kollifinke,I.K. and Pflitzner,U.M.  
TITLE Structure of a heptaubiquitin gene from tomato  
JOURNAL Plant Physiol. 104 (1), 299-300 (1994)  
MEDLINE 94159799

## REFERENCE 2 (bases 1 to 3885)

AUTHORS Pflitzner,U.M.

TITLE Direct Submission

JOURNAL Submitted (03-JUN-1993) U.M. Pflitzner, Botanisches Institut der

LMU, Menzinger Str 67, 8000 Muenchen 19, FRG

FEATURES Location/Qualifiers

source

1..3885

/organism="Lycopersicon esculentum"

/cultivar="craigella"

/db\_xref="taxon:4081"

/clone\_lib="EMBL3"

/clone="1"

75..79

224..234

/note="1"

/rpt\_type=INVERTED

243..252

/note="1"

/rpt\_type=INVERTED

293..302

/note="2"

/rpt\_type=INVERTED

323..334

/note="1"

/rpt\_type=INVERTED

336..345

/note="2"

/rpt\_type=INVERTED

340..350

/note="2"

/rpt\_type=INVERTED

509..516

648..1775

992..1015

/note="1"

/rpt\_type=INVERTED

1050..1067

/note="1"

/rpt\_type=INVERTED

1220..1237

/note="2"

/rpt\_type=INVERTED

1319..1341

/note="2"

/rpt\_type=INVERTED

1557..1561

/note="2"

/rpt\_type=INVERTED

1776..3380

/note="branch point"

/product="ubiquitin"

/protein\_id="CAA51679.1"

/db\_xref="pid:9312160"

/db\_xref="GI:312160"

/db\_xref="SPTREMBL:Q40164"

/translation="MQIFVKTITGKTITLVESSDITDNVAKIQDKGIPPDQORLI

FAGKQLEDGRLADYNIQKESTLHLVLRGMOIFVKITLTKTITLVESSDITDNV

AKIQDKGIPPDQORLI FAGKQLEDGRLADYNIQKESTLHLVLRGMOIFVKITLTK

TGKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLEDGRLADYNIQKEST

LHLVLRGMOIFVKITLTKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGK

QLEDGRLADYNIQKESTLHLVLRGMOIFVKITLTKTITLVESSDITDNVAKI

QDKGIPPDQORLI FAGKQLEDGRLADYNIQKESTLHLVLRGMOIFVKITLTK

ITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLEDGRLADYNIQKESTLHLV

LRGMOIFVKITLTKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLED

GRLADYNIQKESTLHLVLRGMOIFVKITLTKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLED

GRLADYNIQKESTLHLVLRGMOIFVKITLTKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLED

GRLADYNIQKESTLHLVLRGMOIFVKITLTKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLED

GRLADYNIQKESTLHLVLRGMOIFVKITLTKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLED

GRLADYNIQKESTLHLVLRGMOIFVKITLTKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLED

GRLADYNIQKESTLHLVLRGMOIFVKITLTKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLED

GRLADYNIQKESTLHLVLRGMOIFVKITLTKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLED

GRLADYNIQKESTLHLVLRGMOIFVKITLTKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLED

GRLADYNIQKESTLHLVLRGMOIFVKITLTKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLED

GRLADYNIQKESTLHLVLRGMOIFVKITLTKTITLVESSDITDNVAKIQDKGIPPDQORLI FAGKQLED

polyA-signal 3524..3533

BASE COUNT 1106 a 698 c 789 g 1292 t

ORIGIN

Query Match 95.2%; Score 20; DB 7; Length 3885;

Best Local Similarity 100.0%; Pred. No. 42;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgaaagaaagcraaag 20

|||||

DB 78 ATGAAAGAAAGCTAAAG 97

## RESULT 4

CET25C12/c

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

## REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

## FEATURES

source

1..33045

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

/chromosome="X"

/clone="T25C12"

9132..23914

/gene="lin-14"

join(9132..9253,20428..20645,21083..21286,22191..22270,

22753..22929,22980..23088,23135..23350,23398..23563,

23615..23762,23807..23914)

/gene="lin-14"

/note="Embryonic nuclear protein lin-14 form B2; cDNA EST

sequence Z93779.

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

http://webpage.sanger.ac.uk/cgi-bin/display?db=wormaceclass=Sequence&object=T25C12

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

This sequence is the entire insert of clone T25C12. The true right end of clone F48C5 is at 13899 in this sequence. The start of this sequence (1..105) overlaps with the end of sequence Z6107. The end of this sequence (32937..33045) overlaps with the start of sequence Z93779.

2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

Nature 368 (6466), 32-38 (1994)

94150718

Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.

For a graphical representation of this sequence and its analysis see: -

http://webpage.sanger.ac.uk/cgi-bin/display?db=wormaceclass=Sequence&object=T25C12

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

This sequence is the entire insert of clone T25C12. The true right end of clone F48C5 is at 13899 in this sequence. The start of this sequence (1..105) overlaps with the end of sequence Z6107. The end of this sequence (32937..33045) overlaps with the start of sequence Z93779.

Location/Qualifiers

1..33045

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

/chromosome="X"

/clone="T25C12"

9132..23914

/gene="lin-14"

join(9132..9253,20428..20645,21083..21286,22191..22270,

22753..22929,22980..23088,23135..23350,23398..23563,

23615..23762,23807..23914)

/gene="lin-14"

/note="Embryonic nuclear protein lin-14 form B2; cDNA EST



yk260a11.5 comes from this gene"  
 /codon\_start-1  
 /protein\_id="CAA91484.1"  
 /db\_xref="PID:e1349771"  
 /db\_xref="PID:g3880223"  
 /db\_xref="GI:3880223"  
 /db\_xref="SWISS-PROT:Q21446"  
 /translation="MNLQCPPTTVEMCENGNGSTQSVAAOCGCIETLRKVT  
 DLROAVNLILPMLPLPTIGFNATGGLAAPTQAHVYQOGLKRRPVAQPTVPOPE  
 CPQOIRPVLSPPAALONVIMLNPIMGSSSLKPPASPTLNGOIPPTIGETSLQSTDO  
 TVKIGFSSVDNGOKTDSASASAGDNNDIVIDGSESPSSNHSOETAALMSOOT  
 FLNALKDSSEFLTPVPYVERAFLPRLVAPPTNGTNGTARAGPERPRKRVNDIYK  
 IVRNODLSEENISFELIPVKAIASDFFRVSQOIIQOITQCKIEMEVEGCMQ  
 LCKRLAEKRVGPRRLMSQTVAGLNHSYALPKGICYIOHVCRKLYDKFENEDE  
 WDKREARMLAARCRVRHAKTKTHNEEOAEMLSKRGEDPFLNAGLIRPVE  
 ETVSEANILMSDQIKSOLESFAHPRTESETPLEIIQONISLFTHLIRTVESQS  
 PRLGPGP"  
 join(20139..20336,20428..20645,21083..21286,22191..22270,  
 22753..22929,22980..23088,23135..23350,23398..23563,  
 23615..23762,23807..23914)  
 /gene="lin-14"  
 /note="embryonic nuclear protein lin-14 form A"  
 /codon\_start-1  
 /protein\_id="CAA91485.1"  
 /db\_xref="PID:e1349772"  
 /db\_xref="PID:g3880224"  
 /db\_xref="GI:3880224"  
 /db\_xref="SWISS-PROT:Q21446"  
 /translation="MDLPGTSSNDVQPSFVPSPPRPHQAALLPATLAPOLL  
 QPLKSDYLIRSDMSAFKEVTDLRQANLILPMLPLPTIGFNATGGLAAPTQAHV  
 VIQOGLKRRPVAQPTVPOPECPQOIRPVLSPPAALONVIMLNPIMGSSSLKPPAS  
 TLNGOIPPTIGETSLQSTDOQTVKIGFSSVDNGOKTDSASAGDNNDIVIDG  
 SESEPTSSNHSOETAALMSOOTFLNALKDSSEFLTPVPYVERAFLPRLVAPPTNGT  
 GTAAGPERPRKRVNDIYKIVRNODLSEENISFELIPVKAIASDFFRVSQOIIQOITQCK  
 IEMEVEGCMQ LCKRLAEKRVGPRRLMSQTVAGLNHSYALPKGICYIOHVCRKLYDKFENEDE  
 IYVSEANILMSDQIKSOLESFAHPRTESETPLEIIQONISLFTHLIRTVESQS  
 PRLGPGP"  
 join(27443..30141  
 /gene="T25C12.2"  
 join(27443..27598,27646..27774,29771..29932,29986..30141)  
 /gene="T25C12.2"  
 /codon\_start-1  
 /protein\_id="CAA91486.1"  
 /db\_xref="PID:e1349773"  
 /db\_xref="PID:g3880225"  
 /db\_xref="GI:3880225"  
 /db\_xref="SPRKBML:Q22776"  
 /translation="MKTVLFLVAVAVATLVPFEOASAFDCELCRLAVNTDPPVD  
 NELHIVEDKFLAECKKELKIPLEOTCLNASHSEFPIIKELESGTAPEDDITKTVL  
 FLVAVVATLVPFEOASAFDCELCRLAVNTDPPFDEEFHIAEDKFLAECKKELK  
 GIPLEOTCLNASHSEFPIIEMLESGTAPEDDITKTVL  
 join(complement(278543.1:1886..1914),  
 complement(278543.1:1591..1834),  
 complement(278543.1:709..841),  
 complement(278543.1:594..663),  
 complement(278543.1:417..486),  
 complement(278543.1:313..373),  
 complement(278543.1:107..174),  
 complement(293779.1:4276..4438),  
 complement(293779.1:3801..4228),  
 complement(293779.1:3624..3755),  
 complement(293779.1:3300..3578),  
 complement(293779.1:2817..3216),  
 complement(293779.1:2623..2723),  
 complement(293779.1:2483..2574),  
 complement(293779.1:2236..2433),  
 complement(293779.1:1565..2191),  
 complement(293779.1:1366..1518),  
 complement(293779.1:1226..1326),  
 complement(293779.1:1004..1178),  
 complement(293779.1:800..955),  
 complement(293779.1:537..754),

complement(293779.1:110..490),complement(32991..33045),  
 complement(32825..32941),complement(32622..32779),  
 complement(32464..32574),complement(31972..32419),  
 complement(31750..31922),complement(31513..31704),  
 complement(31273..31466),complement(31127..31225),  
 complement(30882..31079),complement(30812..30835))  
 /gene="T25C12.3"  
 /note="similar to EGF-repeats; cDNA EST EMBL:D26687 comes  
 from this gene; cDNA EST EMBL:D26692 comes from this gene;  
 cDNA EST EMBL:D26691 comes from this gene; cDNA EST  
 EMBL:D26694 comes from this gene; cDNA EST  
 EMBL:D26693 comes from this gene; cDNA EST EMBL:D26696 comes from this  
 gene; cDNA EST EMBL:D26695 comes from this gene; cDNA EST  
 EMBL:D26699 comes from this gene; cDNA EST EMBL:D26688  
 comes from this gene; cDNA EST EMBL:D26690 comes from this  
 gene; cDNA EST EMBL:D34849 comes from this gene; cDNA EST  
 EMBL:D2824 comes from this gene; cDNA EST EMBL:D33092  
 comes from this gene; cDNA EST EMBL:D33505 comes from this  
 gene; cDNA EST EMBL:D34204 comes from this gene; cDNA EST  
 EMBL:D34647 comes from this gene; cDNA EST EMBL:D34858  
 comes from this gene; cDNA EST EMBL:D35837 comes from this  
 gene; cDNA EST EMBL:D36857 comes from this gene; cDNA EST  
 EMBL:D67972 comes from this gene; cDNA EST EMBL:D35607  
 comes from this gene; cDNA EST yk383b10.3 comes from this  
 gene; cDNA EST yk383b10.5 comes from this gene; cDNA EST  
 yk259f5.3 comes from this gene; cDNA EST yk259f5.5 comes  
 from this gene; cDNA EST yk295a1.3 comes from this gene;  
 cDNA EST yk295a1.5 comes from this gene; cDNA EST  
 yk294g11.3 comes from this gene; cDNA EST yk294g11.5 comes  
 from this gene; cDNA EST yk285b2.3 comes from this gene;  
 cDNA EST yk285b2.5 comes from this gene; cDNA EST  
 yk276d2.3 comes from this gene; cDNA EST yk276d2.5 comes  
 from this gene; cDNA EST yk243d6.3 comes from this gene;  
 cDNA EST yk243d6.5 comes from this gene; cDNA EST  
 yk239d3.3 comes from this gene; cDNA EST yk239d3.5 comes  
 from this gene; cDNA EST yk226g10.3 comes from this gene;  
 cDNA EST yk226g10.5 comes from this gene; cDNA EST  
 yk240h3.5 comes from this gene"  
 /codon\_start-1  
 /protein\_id="CAA91487.1"  
 /db\_xref="PID:e1349774"  
 /db\_xref="PID:g3880226"  
 /db\_xref="GI:3880226"  
 /translation="MFIOGSCGPFYIPVDSTGNKPTQIOIHVSAGIPYVNTDAK  
 GYVSPSYVADDSYSLNVAADLPSCGGTISIDNOGVTSCIEVSGDSCGKAVG  
 FVSPQSDDPYGSALNGYPMYAAVNDSPRAVOGTTIRQGSALPVRRAATRRY  
 QCGEYTAGMOQCUGNATYTHIDGVANGENRRTIRFSCMOULTISPPATTQAV  
 TTCENGTLLNVAOQGTCCFSELFNKQCEKVNMAAGPDPGNCLECAAGYVGN  
 CODVTCPLNBOYLTNKTILVILRSTVSNOMNATISAIXKELTNDGNGYEVYG  
 FVIVKFNAGVATNTYPAFONAEIAINDSATVAGCCSATFAGISIFNEVALYOK  
 SPVLFPTDAAVDYERKQVILEONTBKPKPIYTRHYPQNSCAEDMOSGOALVYAS  
 YSGLLIRSPVDSIQOIFYIKRATKANSVLLDLSACSTPRYFVDTSTELAI  
 LAGVOGLYSVTDPNGMNTLLKTVDSGTTQLIEISPVVGEHLVTVASVNOQPCSY  
 RVQARSEYDLFISGVNDADSDSEPVQSHIVQGLVAVTVASVNOQPCSY  
 TSNVNVNKRQKPMYISGKRYDSCGADFCDFDSCGADFCDFDSCGADFCDFD  
 ORTTTSGSGPTTPYPPNSGONGVDPFNNAACICPPGPGGOYCNICONGVGR  
 GKMVCGVGTAGTCFCQYCTFENNNDGNNFVYVSTRTMKVAAOKDANO  
 AMTDMQASDKATNKITILVANSNTSYLLVNSRPRDPVAGTVANGNSVYADDT  
 SCQIQIQAMKGAVALLEKRSVAVFDSQPNLSVQLFVDSQVQLNLVAGVS  
 SICTPPNPNOGYTLRLSLQTLGDIYMDLDOIMFLVSMKRSVSHRYVADT  
 SASTYPPNVGWTOSTLTLAVTGDLVVEVTFPGCGQNSDGLVAINPETLNY  
 VAACSGSEFMHROQNCFTPADKLSMLDVTDMCAKAAVALLHDSQNTNYVASQIS

Query Match 87.6%; Score 18.4; DB 36; Length 33045;  
 Best Local Similarity 95.0%; Pred. No. 1.3e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 atgaaaggaagcctaaag 20  
 |||||||  
 Db 1344 atgaaaggaagcctaaag 1325  
 |||||||

RESULT 5

AC004891  
LOCUS AC004891 112621 bp DNA PRI 23-NOV-1998  
DEFINITION Homo sapiens PAC clone DJ0800L12 from 7q34-q36, complete sequence.  
ACCESSION AC004891  
NID 93638952  
VERSION AC004891.1 GI:3638952  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 112621)  
AUTHORS Cotton, M., Tin-Wollam, A., Sutterer, C. and Wilson, C.  
TITLE The sequence of Homo sapiens PAC clone DJ0800L12  
JOURNAL Unpublished (1998)  
REFERENCE 2 (bases 1 to 112621)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
3 (bases 1 to 112621)  
Waterston, R.  
REFERENCE Direct Submission  
AUTHORS Submitted (19-SEP-1998) Department of Genetics, Washington  
JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 19, 1998 this sequence version replaced gi:1213106.  
COMMENT SUBMITTED BY: WUGSC  
Genome Sequencing Center  
Department of Genetics  
Washington University  
St. Louis MO 63108, USA  
mailto://genome.wustl.edu/gsc  
mailto:sapiens@wustl.wustl.edu

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone; and the assembly was  
confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and  
sequencing collaboration between the NHGRI Chromosome 7 Mapping  
Project (Eric D. Green, Director), John D. McPherson in the  
Department of Genetics (Washington University), and the Washington  
University Genome Sequencing Center. For additional information  
about the map position of this sequence, see  
http://www.nhgri.nih.gov/DIR/GRB/CHR7, send  
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc  
mailto:egreen@nhgri.nih.gov.

SOURCE INFORMATION:  
This clone was derived from human PAC library RPCI-4, prepared by  
Pieter de Jong and coworkers at Roswell Park Cancer Institute,  
using the method described by Ioannou et al., Nature Genetics  
6:84-9 (1994). The library is from one male donor. For further  
details, see http://bacpac.med.buffalo.edu/  
The clone is available from Genome Systems, Inc.  
(http://www.genomesystems.com).  
VECTOR: pCYPAC2  
NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the right is DJ0676120, 200 bp overlap.  
Actual start of this clone is at base position 1 of DJ0800L12;  
actual end is at 112621 of DJ0800L12.

FEATURES  
source  
1..112621  
/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q34-q36"  
/clone\_lib="RPCI-4"  
/clone="DJ0800L12"  
369..648  
/rpt\_family="ORSL"  
repeat\_region  
1301..1376  
/rpt\_family="Alu"  
repeat\_region  
1564..1804  
/rpt\_family="L1"  
2106..2455  
/rpt\_family="MALR"  
repeat\_region  
2456..2879  
/rpt\_family="MALR"  
repeat\_region  
2880..2901  
/rpt\_family="MALR"  
repeat\_region  
2909..3003  
/rpt\_family="Other"  
repeat\_region  
3040..3085  
/rpt\_family="(GA)n"  
repeat\_region  
3087..3430  
/rpt\_family="retroviral"  
repeat\_region  
3509..3619  
/rpt\_family="MALR"  
repeat\_region  
3684..3900  
/rpt\_family="MALR"  
repeat\_region  
4209..4377  
/rpt\_family="MALR"  
repeat\_region  
5921..5946  
/rpt\_family="(CA)n"  
repeat\_region  
6788..6834  
/rpt\_family="(CA)n"  
repeat\_region  
7487..7737  
/rpt\_family="MIR"  
repeat\_region  
7687..7751  
/rpt\_family="L2"  
repeat\_region  
9661..9787  
/rpt\_family="Alu"  
repeat\_region  
10159..10464  
/rpt\_family="Alu"  
repeat\_region  
11884..12194  
/rpt\_family="Alu"  
repeat\_region  
12630..13121  
/rpt\_family="retroviral"  
repeat\_region  
13487..13520  
/rpt\_family="AT-rich"  
repeat\_region  
13892..13918  
/rpt\_family="purine-rich"  
repeat\_region  
14150..14425  
/rpt\_family="purine-rich"  
repeat\_region  
14442..14712  
/rpt\_family="Alu"  
repeat\_region  
15139..15333  
/rpt\_family="MIR"  
repeat\_region  
16197..16380  
/rpt\_family="MIR"  
repeat\_region  
16444..16793  
/rpt\_family="MALR"  
repeat\_region  
17422..17728  
/rpt\_family="Alu"  
repeat\_region  
18046..18123  
/rpt\_family="purine-rich"  
repeat\_region  
18708..19190  
/rpt\_family="L1"  
complement(<21462..>75820)  
/gene="WUGSC:DJ0800L12.1"  
complement(join(<21462..21698,46231..46449,75652..>75820))  
/gene="WUGSC:H.DJ0800L12.1"  
/note="contractin-like; similar to U87224 (PID:G1857710);  
H\_DJ0800L12.1"  
/codon\_start=1  
/evidence=not\_experimental

/protein\_id="AAC36298.1"  
/db\_xref="PID:93638953"  
/db\_xref="GI:3638953"  
/translation="NYMNAASFPNPSSYLHFTFOGQTSNDISFYFKTILTPMCFLEN  
MGKEDIKLEKATEVSEFSDVNGPVEIVVSPPLNDODHRYTAERNTVOASIO  
VTRLPODIRKAPRHEGHTRELIEYSOLFVGACGGCGGFLGICRSLRMNGYLIDLEERAKV  
TSGFISGCSGHCTSYCTNCGNGKCLERHGHGSCDSNRYDGTFCNK"  
repeat\_region  
/rpt\_family="MIR"  
22673..42873  
23906..23969  
/rpt\_family="L1"  
repeat\_region  
23970..24260  
/rpt\_family="Alu"  
repeat\_region  
24261..24747  
/rpt\_family="L1"  
repeat\_region  
25328..25902  
/rpt\_family="L1"  
repeat\_region  
26516..26784  
/rpt\_family="Alu"  
repeat\_region  
26785..26804  
/rpt\_family="Alu"  
repeat\_region  
27577..28068  
/rpt\_family="(TAAA)n"  
repeat\_region  
28204..28344  
/rpt\_family="L2"  
repeat\_region  
28512..28888  
/rpt\_family="MIR"  
repeat\_region  
28993..29328  
/rpt\_family="MALR"  
repeat\_region  
29703..29990  
/rpt\_family="Alu"  
repeat\_region  
29991..30014  
/rpt\_family="Alu"  
repeat\_region  
30619..30766  
/rpt\_family="(CAAAA)n"  
repeat\_region  
30828..31224  
/rpt\_family="MIR"  
repeat\_region  
32804..33060  
/rpt\_family="Alu"  
repeat\_region  
32918..33192  
/rpt\_family="L1"  
repeat\_region  
33348..33328  
/rpt\_family="Other"  
repeat\_region  
33509..33676  
/rpt\_family="MER1-type"  
repeat\_region  
33686..33982  
/rpt\_family="Alu"  
repeat\_region  
34331..34442  
/rpt\_family="Alu"  
repeat\_region  
35422..35717  
/rpt\_family="L1"  
repeat\_region  
36469..36797  
/rpt\_family="Alu"  
SRS  
/gene="WUGSC:H.DJ0800L12.1"  
/db\_xref="GI:113778"  
repeat\_region  
36641..36690

Query Match 85.7%: Score 18; DB 11; Length 112621;  
Best Local Similarity 100.0%: Pred. No. 16e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 aaagaagaagctaaagt 21  
Db 78630 AAAAGCAAGCTAAAGT 78647

RESULT 6  
LOCUS HSU66083 73360 bp DNA PRI 29-AUG-1997  
DEFINITION Human contig of two cosmids from LNLX X chromosome library (U93FL,  
U109H10), including MAGE-9 antigen (MAGE9) gene, complete cds, and  
three genes of unknown function.  
ACCESSION U66083  
MD 91519284

VERSION U66083.1 GI:1519284  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM human.  
REFERENCE 1 (bases 1 to 73360)  
AUTHORS Timms,K.M., Bondeeson,M.-L., Ansari-Lari,M.A., Lagerstedt,K.,  
Nelson,D.L., Pettersson,U. and Gibbs,R.A.  
TITLE Molecular and phenotypic variation in patients with severe Hunter  
syndrome  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 73360)  
AUTHORS Timms,K.M., Bondeeson,M.-L., Ansari-Lari,M.A., Lagerstedt,K.,  
Nelson,D.L., Pettersson,U. and Gibbs,R.A.  
TITLE Direct Submission  
JOURNAL Submitted (05-AUG-1996) Department of Human and Molecular Genetics,  
Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030,  
USA

FEATURES  
source Location/Qualifiers  
1..73360  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
precursor\_RNA <1..>2758  
/gene="MAGE9"  
gene /evidence=not\_experimental  
1..2758  
/gene="MAGE9"  
/evidence=not\_experimental  
CDS 1160..2107  
/gene="MAGE9"  
/codon\_start=1  
/product="MAGE-9 antigen"  
/protein\_id="AAB67888.1"  
/db\_xref="PID:91519285"  
/db\_xref="GI:1519285"  
translation="MSLEQRSPHCKRDEDLAEGDGLMGAGOEPTGEEETTSDD  
KEEVSAAAGSSSPQSGASSISVYTLMSQDEGSSQEEEPSSVPAQLEF  
KFOEALKLVAEIVHFLHKRYKREVTAKMLSEYIKYKRYFPVIFKASPFMOVI  
FGTDVXENDPAGHSYLIYALGISDSMTGDSHMPKALILIVIGVITLTKNCAPEE  
VIVHEALSVAGYVYGEKHEMTGEPKRLITDDWQENTLEYRQVPGSDPAHYEFLMGSKA  
HAETSEKVINLVMINAREPICYPSELYEEVEEGEV"

repeat\_region  
complement(5123..5544)  
/rpt\_family="MER1A"  
5773..6002  
/rpt\_family="SVA"  
complement(5775..6008)  
repeat\_region  
/rpt\_family="Alu"  
complement(6949..7587)  
repeat\_region  
/rpt\_family="L1MD2"  
complement(8337..8621)  
repeat\_region  
/rpt\_family="Alu"  
8343..8595  
/rpt\_family="SVA"  
complement(9508..9608)  
repeat\_region  
/rpt\_family="L1"  
9611..9899  
/rpt\_family="L1PA7"  
13634..14034  
repeat\_region  
/rpt\_family="MER28"  
17341..17620  
/rpt\_family="Alu"  
complement(17357..17561)  
repeat\_region  
/rpt\_family="SVA"  
17874..18103  
/rpt\_family="Alu"  
complement(17907..18076)  
repeat\_region  
/rpt\_family="SVA"  
complement(18233..18490)  
repeat\_region  
/rpt\_family="Alu"  
18245..18490  
/rpt\_family="SVA"  
complement(19318..20372)  
repeat\_region

```

repeat_region /rpt_family="MER11A"
complement(21392..21669)
rpt_family="Alu"
repeat_region /rpt_family="Alu"
complement(21716..22106)
rpt_family="L1MB3"
misc_feature 24415..52639
/note="Duplication of bases 178530 - 206055 from contig
formed from cosmids U6986, U142F2, U126G1, U145C10,
1126I0, 169A5 and U84H1"
/note="Region: large scale duplication."
/evidence=experimental
26016..26299
rpt_family="Alu"
complement(26032..26285)
rpt_family="SVA"
repeat_region /rpt_family="SVA"
complement(26695..27058)
rpt_family="MSTa"
complement(26715..28031)
gene /gene="y"
/note="EST from WashU-MERCK EST project; unknown function;
EST clone 46690"
/evidence=experimental
complement(26715..28031)
gene /gene="y"
/note="unknown function"
complement(27172..28489)
rpt_family="THR"
repeat_region /rpt_family="THR"
complement(28712..28934)
rpt_family="MSTa"
31052..31383
rpt_family="MER11A"
complement(37037..43323)
gene /gene="x"
41145..41577
rpt_family="MT1c"
complement(join(<42271..42352,38845..38977,38462,
37037..>37607))
/note="EST from WashU-MERCK EST project; unknown function;
EST clone 115392"
/evidence=experimental
complement(join(42979..43311,42287..>42318))
gene /gene="x"
/note="TA-PCR product; unknown function; clone x'-3a"
/evidence=experimental
complement(join(43030..43322,42271..42318,38845..38977,
38462..38743,37066..>37607))
gene /gene="x"
/note="CDNA from human placental library; unknown
function; clone 191B7"
complement(join(<43030..43261,42271..42318,38845..38977,
38462..38743,37066..>37607))
gene /gene="x"
/note="CDNA from human placental library; unknown
function; clone 48A8"
/evidence=experimental
join(43465..43623,45495..45722,45893..46075,46626..46683,
48487..48598,48854..48936,50694..52183,52914..53011,
53142..53230,65601..65784)
gene /gene="z"
/note="CDNA from human placental library; unknown
function; clone Z'-3-1"
/evidence=experimental
43465..68936
gene /gene="z"
43836..44597
misc_feature /gene="z"
/note="Region with high GC (72%) content; single strand
coverage only; unknown function"
/note="Region: single strand coverage only."
repeat_region /rpt_family="MSR1"
complement(45107..45370)
repeat_region /rpt_family="MSR1"
complement(48476..48744)

```

```

repeat_region /rpt_family="Alu"
48521..48724
rpt_family="SVA"
misc_feature 49536..49920
/gene="z"
/note="Region with high GA (91%) content; single strand
coverage only."
/note="Region: single strand coverage."
join(<53209..53230,65601..>68555)
gene /gene="z"
/note="CDNA from Human B cell (B129) library; unknown
function; clone 161455-2-3"
/evidence=experimental
repeat_region /rpt_family="MER4B"
complement(56174..56748)
rpt_family="MER4B"
repeat_region /rpt_family="Alu"
complement(57171..57475)
rpt_family="Alu"
complement(57669..58064)
rpt_family="L1MB2"
complement(58222..58443)
rpt_family="L1MB1"
58472..58706
rpt_family="SVA"
complement(58473..58722)
rpt_family="Alu"
complement(58747..59276)
rpt_family="L1MB2"
complement(59290..59558)
rpt_family="L1MB2"
complement(59578..59994)
rpt_family="L1MB5"
complement(59990..60323)
rpt_family="Alu"
60051..60257
rpt_family="SVA"
60311..60391
rpt_family="L1PA7"
complement(60402..60731)
rpt_family="L1MA10"
complement(61224..61442)
rpt_family="Alu"
complement(61695..61974)
rpt_family="Alu"
61725..61958
rpt_family="SVA"
complement(62643..63278)
rpt_family="L1MB2"
complement(63305..64041)
repeat_region
Query Match 85.7%; Score 18; DB 11; Length 7360;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 3 gaaaggaagcctaaag 20
|||||
Db 49109 GAAAGGAAGCTAAAG 49092
|||||
RESULT 7
LOCUS XLU19974 2794 bp mRNA VRT 02-FEB-1996
DEFINITION Xenopus laevis poly(A) polymerase type 2 mRNA, complete cds.
ACCESSION U19974
NID 9643645
VERSION U19974.1 GI:643645
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae; Xenopodinae;
Xenopus.
REFERENCE
1 (bases 1 to 2794)
Ballantyne,S., Bilger,A., Astrom,J., Virtanen,A. and Wickers,M.

```

TITLE Poly (A) polymerases in the nucleus and cytoplasm of frog oocytes:  
dynamic changes during oocyte maturation and early development  
JOURNAL RNA 1 (1), 64-78 (1995)  
MEDLINE 96079940  
REFERENCE 2 (bases 1 to 2794)  
AUTHORS Ballantyne,S.M.  
TITLE Direct Submission  
JOURNAL Submitted (18-JAN-1995) Scott M. Ballantyne, Biochemistry,  
University of Wisconsin-Madison, 420 Henry Mall, Madison, WI 53706,  
USA

FEATURES  
source Location/Qualifiers  
1..2794  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/tissue\_type="ovary"  
728..2182  
/codon\_start=1  
/product="poly(A) polymerase type 2"  
/protein\_id="AAC59746.1"  
/db\_xref="PID:g643646"  
/translaton="MLVARTCOLYRNAIATLVKHFVLSKMEPMNPVLLKQPECN  
LNLVMPDPRVNSDRYHLMPIITPAYPOONTYVSVSTAVWVEERKQGLATDEL  
LVKAEKSLDPAHPFQKRYKHYILLASAPTEKORLEWGLVESKIRILVSGEKNE  
ITLAVNPQSPSPSESEKEEFRTMVIGLVFKMESENLSVDLYDIDQSFTDLY  
VDNSMSVSPNTATRTSPNSTGLSGNSPAPVSPVSKKHSVEGVKLVNSDSIDIS  
SGGLNESIPETATPAFSSPTPLTAYVSMPLVNOYOKPVTNTYTKMSPVAGYK  
RTSSPTNEESPKKTKTEDEENDSNSTEDQNTKEPELKEVSEKSSSPVAGALP  
SSORSSTSDLSIDIVLPATPIPVYKNSIKRLNR"

BASE COUNT 860 a 584 c 556 g 794 t

ORIGIN

Query Match 84.8% Score 17.8 DB 4: Length 2794;  
Best Local Similarity 90.5% Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgaaagaagaactaaagt 21  
|||||||  
Db 2578 ATAAAGCAATCTAAAGT 2558

RESULT 8  
SC9959 40397 bp DNA PLN 11-AUG-1997  
LOCUS S.cerevisiae chromosome XIII cosmid 9959.  
ACCESSION Z49938 Z71257  
NID 9887599  
VERSION Z49939.1 GI:887599  
KEYWORDS dHvdfolate reductase: END1; ERG8; MRE11; MRPL44; MTF1;  
phosphomevalonate kinase; ribonuclease H; ribosomal protein L44;  
RNH1; transfer RNA-Arg; ubiquitin carboxyl-terminal hydrolase.  
baker's yeast.  
SOURCE Saccharomyces cerevisiae  
ORGANISM Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
Saccharomycetaceae; Saccharomycetes.  
REFERENCE 1 (bases 1 to 40397)  
AUTHORS Skelton,J. and Churcher,C.M.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 40397)  
AUTHORS Barrell,B., Rajandream,M.A. and Walsh,S.V.  
TITLE Direct Submission  
JOURNAL Submitted (22-JUN-1995) Saccharomyces cerevisiae chromosome XIII  
sequencing Project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge  
CB10 1RQ E-mail: barrell@sanger.ac.uk  
3 (bases 16981 to 22170)  
Venema,J. and Tollervey,D.  
REFERENCE RRP5 is required for formation of both 18S and 5.8S rRNA in yeast  
JOURNAL EMBO J. 15 (20), 5701-5714 (1996)  
MEDLINE 97051828  
COMMENT Notes:  
All CDS over 100 codons have been analysed. CDS that are completely

overlapped and those that are overlapped by more than 50%  
of their length by a larger CDS have been omitted from this  
analysis.  
Details of the omitted CDS are available on request. The more  
significant matches with motifs in the PROSITE database are  
also included but some of these may be fortuitous. The length in  
codons and the calculated codon adaptation index (CAI)  
is given for each CDS.  
Cosmid 9959 overlapped at 5' by cosmid 8261, emb1 entry SC8261X,  
accessionno. Z49809 and at the 3' by cosmid 9408, emb1 entry SC9408,  
accession no. Z48756.

FEATURES  
source Location/Qualifiers  
1..40397  
/organism="Saccharomyces cerevisiae"  
/strain="AB972"  
/db\_xref="taxon:4932"  
/chromosome="XIII"  
/clone="cosmid 9959"  
/map="13R"  
<1..3156  
/note="Ym9959.01, len: > 1051, CAI: 0.15, incomplete  
unknown orf, overlaps and extends cosmid 8261 orf  
Ym261.13"

CDS  
/codon\_start=1  
/product="unknown"  
/protein\_id="CA90190.1"  
/db\_xref="PID:9887600"  
/db\_xref="GI:887600"  
/db\_xref="SWISS-PROT:Q03661"  
/translation="SSSVELEVERKSKKLDGSTEKELVPLSDITTTINNSLGNDSI  
YSLDDADAISEMLTDVPLMEIKTTPREYVIVSESVSTSYEDNTVAMPQVEYSP  
FMNDPENSINDDEKHKDLKSTLALAPAFKDAEFAVGATKSLTSGHTNF  
HTSKETKQVSDLDSESTENFTEFNENTGDEKNSKNPVGANSTDKSTENTKEYES  
AIYNTVNTGSSCEDIEETASNYENRNYCEKMEENAEKSSGCDYQNDQSKTOS  
FSTSDSPNRPESNDNTEFSTKTKVRNSDLEDEBSLKEETKAEVYDKLDEESDSY  
EQDYADPEPNDGSSNENIVKGRKDTLGLVEEENKVNKVEETLEFANVSSVAV  
QNDMDHVDYNOQAQNYEAGERRKYIIONTDEBAHSIIERIDENAGNMIEPENS  
CVERTHNEVLEFERRATTIENTKALENTNHHQDOVSOACSDSDOSTAKENEGSAK  
HNDIRVSSSEISEVEPLKESDRSNISFSPRIVIGAVYGVGVAVDVAESFVKIDY  
MDSQDNDVNDIGYNDIPKNSSTASVAMKGVSSKERSDDEAVITIGGVYAEAHN  
DNGNSRVYINDPTNGAYIEDESEVRQOYKDKLNLKSEPLEVEGLSQHREKDKH  
SENEEFDTLYIGDITSANIHNSNAPDDIKRQQLKLSLDENYSQRLLEDSRKNDK  
SDEVNTSREDLTFERKSVNEKYAGALIEDFSLDISIQHREHEDLDSNNBERSY  
ELNSEPEEALYLEIEGPEPTAASSKMNDEDRQNIPTSDPSPDKEEYTDY  
PYNSENSHITAEKSAPTSPEYELIFSDTPVPMELINDEIPTLEKDKTNVSLDD  
RSEHLSHDVDNPHDINSITIKYNEGEPHQAVDIPKAYKVEEODEMKSYLEEQ  
KPSWELINDSSPENNDETNEKDKTKAKKSRKNINSRRKRKITEGSSASNT  
KRRGRHPKSRGQNTHPYVDK"

1..312  
/note="overlapp with cosmid 8261, EMBL SC8261X, Z49809,  
positions 27890 to 28201"  
3363..4718  
/gene="ERG8"  
/gene="ERG8"  
/db\_xref="SGD:S0004833"  
3363..4718  
/gene="ERG8"  
/note="Ym9959.02, ERG8 gene, len: 451, CAI: 0.15,  
SW:ERG8\_YEAST\_P24521, phosphomevalonate kinase, contains  
P500627 GMP kinases putative ATP-binding domain, conflict  
at C-terminus due to a frameshift with ERG8\_YEAST"  
/codon\_start=1  
/product="ERG8p"  
/protein\_id="CA980191.1"  
/db\_xref="PID:9887601"  
/db\_xref="GI:887601"  
/db\_xref="SWISS-PROT:P24521"  
/translation="MSLRAFSAPGKILAGYLVDTKYEAFVYGLSARHAAVHPY  
GSIQSGKPEFVYRKQKDFGLYHISPGTIPVSGSKSPTEKYANVPSFK  
PMNDICNRRLPYIDIFSDDAYHSQEDSVTEHGNRRLSHRIEPLVPTGLGSSHG  
LVTVLTALASFFVSDLENNVDKRYREIVHNLAVQACQAGKIGSGEDVAAVAGSIR  
YRREPPLISNLDIGSATYGSKLALHYDEEDNITIKSNHLPSGLTLWMDGIKNGSE  
TVRLVQVKVMYDSHPESIKITYELDLSARMDGLSKLDRLHETHHDYSDQIFPSL

misc\_feature  
 ERNDCTQKYPITTEVDAVATIRRSFKITRESGADIEPPVQISLDDCOTLKGLT  
 CLIPAGGDAIAVITKODVDLRAQTANDKRSKQVMDLVTDGAVKREKDEPYLLD  
 K\*  
 /gene="ERG8"  
 /note="P500627 GMP kinases putative ATP-binding domain"  
 /db\_xref="SGD:S000483"  
 /complement(4978..6492)  
 /note="YM9959.03C, unknown, len: 504, CAI: 0.15, possible  
 membrane protein"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="CA90192.1"  
 /db\_xref="PID:9887602"  
 /db\_xref="GI:887602"  
 /db\_xref="SWISS-PROT:004991"  
 /translation="MTSTRTLEAGVACACWICLPASGIIIFGFAALKPILLSEGYHE  
 LCDPDDGRLCTADADKINFIPLASATYNTMALPYGKIDMSPRCGIIIGSLP  
 LASGNFISAKHLVSLMDPTLVGTTLLAAGPVFTSCFOLANSFQSGTVIALTGS  
 FDSSALFLYLLYQMPPTLVSRFFLLIIVFLLACQLTIMPHSKTVNHA  
 KIVAGLDENGRLIGDNGSGIIPDOEROSLIAEREDSIPSRPQRKSVLEYVE  
 DKLOKSGGIFGVLRKSAVEQIKSPWFYLLMFLVAMLRINTFYATVROEYLLN  
 DDPLAKINSIFDMLPLGGAVSIPFGILLHPTDLSITLITFTTAIGVEGLIPN  
 SEFMNLIGVLLVYRPPYTVSDYSKVRFPDRTGYGLSCICGIFMNSQILD  
 KMTHTFNANPPINLLVITLVFVSLITFTYIKSQLLPKPVNERGLSSNTQTI"  
 /complement(6685..7356)  
 /note="YM9959.04C, unknown, len: 223, CAI: 0.10, similar  
 to SW:DR\_SCHPO P36591,dihydrofolate reductase (29.4%  
 identity in 218 aa overlap)"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="CA90193.1"  
 /db\_xref="PID:9887603"  
 /db\_xref="GI:887603"  
 /db\_xref="SWISS-PROT:005015"  
 /translation="MTKNYLMHGLAOSDYRASKTKRAMEIKYKLYTPAPNE  
 FPPADVPDFLEVIADAPGDEGNTCYLALENDPSTGFTITOTTIDILHNTYENCP  
 FAGIVFSQAGVAGLATFENGLGLTTEEPLEFFMAVSGFFRPQOQOEYDLH  
 PISVPSLHVQGLDITTEPAKVOGLYNSCTEDSRLLMHSGHFVPSNRGFEVRYAOM  
 LQOLT"  
 /complement(762..917)  
 /note="YM9959.05, unknown, len: 471, CAI: 0.17, similar to  
 putative ubiquitin carboxyl-terminal hydrolases, eg  
 SW:UBPX\_HUMAN, P40818, Probable ubiquitin  
 carboxyl-terminal hydrolase (31.5% identity in 352 aa  
 overlap); contains P500973 Ubiquitin carboxyl-terminal  
 hydrolases family 2 signature 2"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="CA90194.1"  
 /db\_xref="PID:9887604"  
 /db\_xref="GI:887604"  
 /db\_xref="SWISS-PROT:P50102"  
 /translation="MSICPHIOVFONESKQGVLTCAATYIINHSPKREKLTAM  
 KGTCHETNSGATFMCLOCGFCGNNHSHFSKSHQKIGTETINSMSLLOCLINPY  
 ICNIDILINAILAKTWDCYCTWPMSPMRKGLSGLLNMGSTGNSGSLLOCLINPY  
 FLRHSQSIHNSNCKVRSPDKCFSCALDVIHVELYALNTKQASSSTINQCTFI  
 LITCAKINQNLNAGYSDQDAHEFWQFIINQIHSYVLDLPNKEYSRANKQCEIVH  
 TVEGSLSEIYVPGQNNSTTIDPFLSLDKDKRKLKIDLSFRRKQCEIVH  
 HGCENSTODAKOIGIKHLPISVLYQLKRFPHLNGSRKRLDDITEPFTYINMAYNC  
 STREKDNHSEKRVDPDIYELIGIVSHKGTVEGHIAICKISGQWFEFDSMVSSI  
 SOEEVLEKQAVLYLFTIKOVN"  
 /complement(9045..9922)  
 /note="P500973 Ubiquitin carboxyl-terminal hydrolases  
 family 2 signature 2"  
 /complement(9622..11700)  
 /gene="MRE11"  
 /db\_xref="SGD:S0004837"  
 /complement(9622..11700)  
 /gene="MRE11"  
 /note="YM9959.06C, MRE11 gene, len: 692, CAI: 0.14,  
 SW:MRL1\_YEAST, P32829, required for repair and meiotic  
 recombination, conflict, this sequence is longer due to a

CDS  
 ERNDCTQKYPITTEVDAVATIRRSFKITRESGADIEPPVQISLDDCOTLKGLT  
 CLIPAGGDAIAVITKODVDLRAQTANDKRSKQVMDLVTDGAVKREKDEPYLLD  
 K\*  
 /gene="ERG8"  
 /note="P500627 GMP kinases putative ATP-binding domain"  
 /db\_xref="SGD:S000483"  
 /complement(4978..6492)  
 /note="YM9959.03C, unknown, len: 504, CAI: 0.15, possible  
 membrane protein"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="CA90192.1"  
 /db\_xref="PID:9887602"  
 /db\_xref="GI:887602"  
 /db\_xref="SWISS-PROT:004991"  
 /translation="MTSTRTLEAGVACACWICLPASGIIIFGFAALKPILLSEGYHE  
 LCDPDDGRLCTADADKINFIPLASATYNTMALPYGKIDMSPRCGIIIGSLP  
 LASGNFISAKHLVSLMDPTLVGTTLLAAGPVFTSCFOLANSFQSGTVIALTGS  
 FDSSALFLYLLYQMPPTLVSRFFLLIIVFLLACQLTIMPHSKTVNHA  
 KIVAGLDENGRLIGDNGSGIIPDOEROSLIAEREDSIPSRPQRKSVLEYVE  
 DKLOKSGGIFGVLRKSAVEQIKSPWFYLLMFLVAMLRINTFYATVROEYLLN  
 DDPLAKINSIFDMLPLGGAVSIPFGILLHPTDLSITLITFTTAIGVEGLIPN  
 SEFMNLIGVLLVYRPPYTVSDYSKVRFPDRTGYGLSCICGIFMNSQILD  
 KMTHTFNANPPINLLVITLVFVSLITFTYIKSQLLPKPVNERGLSSNTQTI"  
 /complement(6685..7356)  
 /note="YM9959.04C, unknown, len: 223, CAI: 0.10, similar  
 to SW:DR\_SCHPO P36591,dihydrofolate reductase (29.4%  
 identity in 218 aa overlap)"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="CA90193.1"  
 /db\_xref="PID:9887603"  
 /db\_xref="GI:887603"  
 /db\_xref="SWISS-PROT:005015"  
 /translation="MTKNYLMHGLAOSDYRASKTKRAMEIKYKLYTPAPNE  
 FPPADVPDFLEVIADAPGDEGNTCYLALENDPSTGFTITOTTIDILHNTYENCP  
 FAGIVFSQAGVAGLATFENGLGLTTEEPLEFFMAVSGFFRPQOQOEYDLH  
 PISVPSLHVQGLDITTEPAKVOGLYNSCTEDSRLLMHSGHFVPSNRGFEVRYAOM  
 LQOLT"  
 /complement(762..917)  
 /note="YM9959.05, unknown, len: 471, CAI: 0.17, similar to  
 putative ubiquitin carboxyl-terminal hydrolases, eg  
 SW:UBPX\_HUMAN, P40818, Probable ubiquitin  
 carboxyl-terminal hydrolase (31.5% identity in 352 aa  
 overlap); contains P500973 Ubiquitin carboxyl-terminal  
 hydrolases family 2 signature 2"  
 /codon\_start=1  
 /product="unknown"  
 /protein\_id="CA90194.1"  
 /db\_xref="PID:9887604"  
 /db\_xref="GI:887604"  
 /db\_xref="SWISS-PROT:P50102"  
 /translation="MSICPHIOVFONESKQGVLTCAATYIINHSPKREKLTAM  
 KGTCHETNSGATFMCLOCGFCGNNHSHFSKSHQKIGTETINSMSLLOCLINPY  
 ICNIDILINAILAKTWDCYCTWPMSPMRKGLSGLLNMGSTGNSGSLLOCLINPY  
 FLRHSQSIHNSNCKVRSPDKCFSCALDVIHVELYALNTKQASSSTINQCTFI  
 LITCAKINQNLNAGYSDQDAHEFWQFIINQIHSYVLDLPNKEYSRANKQCEIVH  
 TVEGSLSEIYVPGQNNSTTIDPFLSLDKDKRKLKIDLSFRRKQCEIVH  
 HGCENSTODAKOIGIKHLPISVLYQLKRFPHLNGSRKRLDDITEPFTYINMAYNC  
 STREKDNHSEKRVDPDIYELIGIVSHKGTVEGHIAICKISGQWFEFDSMVSSI  
 SOEEVLEKQAVLYLFTIKOVN"  
 /complement(9045..9922)  
 /note="P500973 Ubiquitin carboxyl-terminal hydrolases  
 family 2 signature 2"  
 /complement(9622..11700)  
 /gene="MRE11"  
 /db\_xref="SGD:S0004837"  
 /complement(9622..11700)  
 /gene="MRE11"  
 /note="YM9959.06C, MRE11 gene, len: 692, CAI: 0.14,  
 SW:MRL1\_YEAST, P32829, required for repair and meiotic  
 recombination, conflict, this sequence is longer due to a

frameshift with SW:MRL1\_YEAST"  
 /codon\_start=1  
 /db\_xref="SGD:S0004837"  
 /product="MRE11"  
 /protein\_id="CA90195.1"  
 /db\_xref="PID:9887605"  
 /db\_xref="GI:887605"  
 /db\_xref="SWISS-PROT:P32829"  
 /translation="MDYDPDPTIRILITTTNNHGVNENDPITGDDSKTFHEVYMLAK  
 NNDYDVGSGDLFEHVNKPSKRSKYOVILKRLCCMGDKCELELLSDPSQVHFDEF  
 TNVYEDNPENISIPVFGISGNHDAAGDSILCPMDILHATGILINHGKVIENSKIV  
 VPLPFGKSTFLALYGLAARDELEPTFPDGCYTFEVPNNRBEENMLACYNHNG  
 HNTAFLEPDLPLDMLVINGHECTPMLVHPINFLNFDVLPQSSVATSLCAEQ  
 PKYVILDEIKTGEAPKMTPIPLETIRFKKMSISQVPHLRPHDKATSKYLEQVE  
 EMIYANDETKOKLADDEGDMVALEKPLIRLVVDYASASNTQSPIDYGVENRRFS  
 NRPFGVANGNNVQVFKRSPVTRSGSGNGTISDRVEXKLFSESGELEVOYLV  
 NDILNKQSLILPVGILNEAVKRVDEDTAKKEFTISHETNSNVGLISTNEFLRV  
 DAEMKALIKRYKANSVRTPPRENDETNAFNGNLDPSFRSNNRVRIGSPDITDS  
 HVDNESRTHTSQAESKRPYSKPRVTRATKRIIPASDSTVISDAENEJGDNDAOD  
 DVDIDENDILMVSTDEBDASVGLNGRRTKTRPAASRTASNRGRASRPKTDI  
 LGSILAKRK"  
 /complement(12007..12450)  
 /gene="MRP14"  
 /db\_xref="SGD:S0004838"  
 /complement(join(12007..12245,12393..12450))  
 /gene="MRP14"

Query Match  
 Best Local Similarity 84.8% Score 17.8; DB 7; Length 40397;  
 90.5%; Pred. No. 2.2e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atgaaaggaagctaaagt 21  
 1 ||||| ||||| |||||  
 Db 30511 AGGAAAGGAAGTAAAGT 30531

RESULT 9  
 HSJ753D10  
 LOCUS  
 DEFINITION  
 Human DNA sequence from clone 753D10 on chromosome 20, complete  
 sequence.  
 ACCESSION  
 AL049651  
 NID  
 94741619  
 VERSION  
 AL049651.2 GI:4741619  
 KEYWORDS  
 HTG; CPG Island.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE  
 1 (bases 1 to 97912)  
 AUTHORS  
 Matthews, L.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (09-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone  
 requests: clonerequests@sanger.ac.uk  
 On May 4, 1999 this sequence version replaced gi:4678569.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em, EMBL; Sw, SWISSPROT; Tr, TRNBL  
 This sequence is the entire insert of clone 753D10. This sequence  
 has been finished according to sequence map criteria as follows. An  
 attempt is made to resolve all sequencing problems, such as  
 compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.  
 This sequence was generated from part of bacterial clone contigs of

human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>  
753D10 is from the library RPCI4 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTON:PCIRPAC2>.

## FEATURES

Source

1..97912  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/clone="753D10"  
/clone\_id="RPCI4"  
1..2243  
/note="L1PB3 repeat: matches 3894..6146 of consensus"  
2246..2515  
/note="AluX repeat: matches 3..289 of consensus"  
2516..2749  
/note="L1M4 repeat: matches 3613..3847 of consensus"  
2750..3311  
/note="L1R26 repeat: matches 1..603 of consensus"  
3312..3392  
/note="L1M4 repeat: matches 3533..3613 of consensus"  
4908..5307  
/note="Charlie4a repeat: matches 49..493 of consensus"  
6793..6862  
/note="L1R1A1 repeat: matches 61..138 of consensus"  
6863..7227  
/note="L1M4 repeat: matches 3626..4016 of consensus"  
7226..7478  
/note="L1MD repeat: matches 350..598 of consensus"  
7479..8086  
/note="L1PB3 repeat: matches 5532..6147 of consensus"  
8087..8418  
/note="L1MD repeat: matches 4..350 of consensus"  
8435..8926  
/note="HAL1 repeat: matches 274..810 of consensus"  
9028..9081  
/note="L1R1A1 repeat: matches 106..154 of consensus"  
9082..9392  
/note="AluX repeat: matches 1..312 of consensus"  
9393..9575  
/note="L1R1A1 repeat: matches 154..362 of consensus"  
9905..10360  
/note="L1R3 repeat: matches 1..513 of consensus"  
10452..10849  
/note="L1R1B repeat: matches 1..390 of consensus"  
11444..11807  
/note="THE1C repeat: matches 1..371 of consensus"  
12305..12549  
/note="L1MC3 repeat: matches 7491..7739 of consensus"  
12564..12720  
/note="FRAM repeat: matches 1..157 of consensus"  
12724..12950  
/note="L1MC3 repeat: matches 7247..7485 of consensus"  
14796..15096  
/note="AluY repeat: matches 1..301 of consensus"  
15331..15520  
/note="MER53 repeat: matches 1..189 of consensus"  
15978..16215  
/note="AluJ repeat: matches 2..291 of consensus"  
16643..22761  
/note="L1PA2 repeat: matches 2..6144 of consensus"  
23043..23087  
/note="L2 repeat: matches 2646..2695 of consensus"  
23210..23287  
/note="2 copies 39 mer 9% conserved"  
26463..26633  
/note="MER5A repeat: matches 1..189 of consensus"  
26711..26742  
/note="8 copies 4 mer gtgt 100% conserved"  
26711..26744  
/note="17 copies 2 mer gt 100% conserved"

27029..27114  
/note="L2 repeat: matches 2628..2709 of consensus"  
29111..29160  
/note="MIR repeat: matches 113..159 of consensus"  
29161..29517  
/note="L1R1A1 repeat: matches 3..365 of consensus"  
29518..29548  
/note="MIR repeat: matches 159..190 of consensus"  
30318..30526  
/note="L1R16C repeat: matches 40..241 of consensus"  
30535..30676  
/note="L1R16C repeat: matches 214..356 of consensus"  
31837..32231  
/note="L1R2CB repeat: matches 130..501 of consensus"  
34206..34433  
/note="L1M4 repeat: matches 4882..5135 of consensus"  
34434..34679  
/note="MER46a repeat: matches 5..236 of consensus"  
34680..34794  
/note="L1M4 repeat: matches 5135..5243 of consensus"  
34850..35115  
/note="MIR repeat: matches 15..262 of consensus"  
35119..35234  
/note="29 copies 4 mer catc 67% conserved"  
35516..35845  
/note="AluY repeat: matches 1..304 of consensus"  
36129..36418  
/note="AluY repeat: matches 1..288 of consensus"  
37096..37162  
/note="MER5B repeat: matches 106..173 of consensus"  
38039..38354  
/note="79 copies 4 mer tata 76% conserved"  
38039..38352  
/note="157 copies 2 mer ta 76% conserved"  
38055..38354  
/note="10 copies 30 mer 77% conserved"  
38355..38410  
/note="14 copies 4 mer atat 79% conserved"  
38355..38408  
/note="27 copies 2 mer at 80% conserved"  
39756..40043  
/note="AluX repeat: matches 24..311 of consensus"  
42556..42693  
/note="L2 repeat: matches 2555..2689 of consensus"  
42866..42992  
/note="AluSg/X repeat: matches 1..127 of consensus"  
43031..43320  
/note="AluSg repeat: matches 1..302 of consensus"  
43323..43475  
/note="L2 repeat: matches 2246..2399 of consensus"  
43560..43863  
/note="AluSg repeat: matches 1..304 of consensus"  
44216..44702  
/note="MER66B repeat: matches 1..486 of consensus"  
45120..45317  
/note="MIR repeat: matches 2..189 of consensus"  
45549..45879  
/note="L1MC repeat: matches 1073..1400 of consensus"  
45957..46367  
/note="L1MC repeat: matches 549..967 of consensus"  
46397..46572  
/note="MIR repeat: matches 19..192 of consensus"  
47319..47569  
/note="MIR repeat: matches 19..262 of consensus"  
47577..47786  
/note="7 copies 30 mer 72% conserved"  
47579..47779  
/note="67 copies 3 mer tac 71% conserved"  
47582..47776  
/note="5 copies 39 mer 79% conserved"  
47869..48078  
/note="MER53 repeat: matches 3..189 of consensus"  
48135..48162

repeat\_region /note="14 copies 2 mer ac 96% conserved"  
49469..50093  
repeat\_region /note="MER41B repeat: matches 1..635 of consensus"  
51836..52264  
repeat\_region /note="MER55C repeat: matches 1..461 of consensus"  
52814..52966  
repeat\_region /note="LTR26 repeat: matches 308..461 of consensus"  
52967..53373  
repeat\_region /note="MER48 repeat: matches 1..398 of consensus"  
53374..53459  
repeat\_region /note="LTR26 repeat: matches 461..546 of consensus"  
53637..53921  
repeat\_region /note="MER49 repeat: matches 448..736 of consensus"  
54325..54540  
repeat\_region /note="LIMC4 repeat: matches 7706..7971 of consensus"  
54541..54927  
repeat\_region /note="MER48 repeat: matches 1..398 of consensus"  
54928..55004  
repeat\_region /note="LIMC4 repeat: matches 7640..7706 of consensus"

Query Match 84.8%; Score 17.8; DB 10; Length 97912;  
Best Local Similarity 90.5%; Pred. No. 2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 atgaaaggaagctaaagt 21  
|||||  
Db 84740 ATGAAAGCAATTAAGT 84760

RESULT 10  
LOCUS HS1114A1 153361 bp DNA HTG 11-JUN-1999  
DEFINITION Homo sapiens chromosome 20 clone 1114A1, WORKING DRAFT SEQUENCE, in  
unordered pieces.  
ACCESSION AL0355684  
NID 95051830  
VERSION AL0355684.18 GI:5051830  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT Direct Submission  
Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk  
On Jun 12, 1999 this sequence version replaced gi:5050930.  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in progress  
and the release of this data is based on the understanding that the  
sequence may change as work continues. The sequence may be  
contaminated with foreign sequence from E.coli, yeast, vector,  
phage etc. Order of segments is not known; 800 n's separate  
segments. Unfinished: dj1114A1 Contig\_ID: 01633 acc=AL0355684  
length: 1068 bp unfinished: dj1114A1 Contig\_ID: 03102  
acc=AL0355684 length: 1703 bp unfinished: dj1114A1 Contig\_ID:  
02484 acc=AL0355684 length: 131637 bp unfinished: dj1114A1  
Contig\_ID: 01371 acc=AL035684 length: 1223 bp.  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES  
Source  
1..153361  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/clone="1114A1"

BASE COUNT 36984 a 34517 c 36452 g 43008 t 2400 others  
ORIGIN

Query Match 84.8%; Score 17.8; DB 34; Length 153361;  
Best Local Similarity 90.5%; Pred. No. 1.8e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 atgaaaggaagctaaagt 21  
|||||  
Db 107132 ATGAAAGCAAGCTAATAGT 107112

RESULT 11  
LOCUS HS232N11 111641 bp DNA HTG 11-JUN-1999  
DEFINITION Homo sapiens chromosome 20 clone 232N11, WORKING DRAFT SEQUENCE, in  
unordered pieces.  
ACCESSION AL031656  
NID 95050948  
VERSION AL031656.9 GI:5050948  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT Direct Submission  
Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk  
On Jun 11, 1999 this sequence version replaced gi:4995276.  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in progress  
and the release of this data is based on the understanding that the  
sequence may change as work continues. The sequence may be  
contaminated with foreign sequence from E.coli, yeast, vector,  
phage etc. Order of segments is not known; 800 n's separate  
segments. Unfinished: dj232N11 Contig\_ID: 00640 acc=AL031656  
length: 111641 bp.  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES  
source  
1..111641  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/clone="232N11"

BASE COUNT 32992 a 24273 c 22868 g 31508 t  
ORIGIN

Query Match 84.8%; Score 17.8; DB 34; Length 111641;  
Best Local Similarity 90.5%; Pred. No. 1.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 atgaaaggaagctaaagt 21  
|||||  
Db 16306 ATGAAAGCAAGCTTAAGT 16326

RESULT 12  
LOCUS ATE24G24/c 99856 bp DNA PIN 25-MAR-1999  
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F24G24 (RSSA  
project).  
ACCESSION AL049488  
NID 94538949  
VERSION AL049488.1 GI:4538949  
KEYWORDS thale cress.  
SOURCE Arabidopsis thaliana  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;



euophyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Caprales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 99856)  
Bevan, M., Murphy, G., Riddle, P., Hudson, S., Bancroft, I., Mewes, H.W., Mayer, K.F.X. and Schueller, C.  
Unpublished  
2 (bases 1 to 99856)  
EU Arabidopsids sequencing project.  
Direct Submission  
Submitted (23-MAR-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk  
Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosome 4 can be viewed at: <http://webvr.mips.biochem.mpg.de/proj/thal/>.  
Location/Qualifiers  
1. .99856  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="4"  
1. .4596  
/note="overlap to BAC F28M11; please refer to this entry for analysis and annotation"  
join(2780..2971,3173..3265,3758..3991,4498..4657,4909..5027,5489..5818)  
/gene="F24G24.10"  
/note="similarity to IB1C3-1 protein, Arabidopsis thaliana, AJ011845"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB39774.1"  
/db\_xref="PID:e1420109"  
/db\_xref="PID:g4538950"  
/db\_xref="GI:4538950"  
/translation="MKMKLGYSLPYVMCMILFYCIILCSILMSHGVEAKALKRNE DLIEOKLEILINKHTVLIKICNGERGVCDYKOPDLDSIKNRHFHKMLMSPE EGSKIKQTHINKTFGFMKNGVRPRTGTPILVLSKALSKSPDGSNDPOSSVS KIKYKPTSSNGGHHFAVVRTKGRPRRNGVAMINSPVGPMEFSAQSMHQIGNE FVOVGWTVRGHCYNPACPVGSGIILVSHVTPGLIKHNDFFELSIKDKIYHMMWL MGNSSSTWKEIGFEMFPRKESGTEGVNGEYSPATSPPMGNSHFPKSGPKIDS YVRLITWDENYGLDMYVKNTERRSNVCYVKAQSRFGQM"  
2780..2971  
/gene="F24G24.10"  
/number=1  
2780..5818  
/gene="F24G24.10"  
2972..3172  
/gene="F24G24.10"  
/number=1  
3173..3265  
/gene="F24G24.10"  
/number=2  
3266..3757  
/gene="F24G24.10"  
/number=2  
3758..3991  
/gene="F24G24.10"  
/number=3  
3992..4498  
/gene="F24G24.10"  
/number=3  
4499..4657  
/gene="F24G24.10"  
/number=4  
4658..4908  
/gene="F24G24.10"  
/number=4  
4909..5027  
/gene="F24G24.10"

/number=5  
5028..5488  
/gene="F24G24.10"  
/number=5  
5489..5818  
/gene="F24G24.10"  
/number=6  
5819..6885  
/gene="F24G24.20"  
/number=1  
6886..10641  
/gene="F24G24.20"  
join(6885..8915,9402..9554,10285..10641)  
/gene="F24G24.20"  
/note="similarity to IB1C3-1 protein, Arabidopsis thaliana, AJ011845"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB39775.1"  
/db\_xref="PID:e1420110"  
/db\_xref="PID:g4538951"  
/db\_xref="GI:4538951"  
/translation="MRMSYSEGSKKRKRTHRTTSGHEWNGVCCITGVPIPRVTKD ALRKMSFSDNSNPQSSWSKTYKPASSIDDHFAVVRTKGRSYNGASMNINTEP SVGPMPQFSASRMHFQIGNEFIQVGMIDKINGMNLMTSWEVEVGPSSRFESSGT MYMGGEVYSPSPNPMPGNSHYPKSGPYDVSIVRLITVDENYNTDKTVKNTERYSD SCYKVDATETEFMSHVGLIYIGPGCK"  
8916..9401  
/gene="F24G24.20"  
/number=1  
9402..9554  
/gene="F24G24.20"  
/number=2  
9555..10284  
/gene="F24G24.20"  
/number=2  
10285..10641  
/gene="F24G24.20"  
/number=3  
11823..12747  
/gene="F24G24.30"  
join(11823..11915,12019..12747)  
/gene="F24G24.30"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB39776.1"  
/db\_xref="PID:e1420111"  
/db\_xref="PID:g4538952"  
/db\_xref="GI:4538952"  
/translation="MTSIIISAVSKLGRGVNQTENKRNPGASEFDIINQAKMLITQYEX YKRGKFPDHYWPIIKIGIEKFAVNDNMKTPPAFOEGADVSSSFSSINTESSPEPGMS IDLNMDSEDNFTLSRPPGLKRAKRQOSEDFKQLEQNDKLIATIKTSERNETI OROKIIVARKKEENKILIFADLNSISDPSSAAVYENKRIETLEKRAQTINQHEDEGQSO YHGSQYRASHYQESLFLFKOVQGEPPQGEDKRSPPNQDEFTQYNYLSCIGNNFP"  
11823..11915  
/gene="F24G24.30"  
/number=1  
11916..12018  
/gene="F24G24.30"  
/number=1  
12019..12747  
/gene="F24G24.30"  
/number=2  
12748..14529  
/gene="F24G24.40"  
/number=2  
14530..15088  
/gene="F24G24.40"  
complement(14529..14816)  
/gene="F24G24.40"  
/number=1  
complement(join(14529..14816,14888..15088))  
/gene="F24G24.40"  
/note="strong similarity to zinc-finger protein R2931, Oryza sativa, PIR3:JE0116"  
/codon\_start=1

/product="zinc-finger-like protein"  
/protein\_id="CAB39777.1"  
/db\_xref="PID:e1420112"  
/db\_xref="PID:g4538953"  
/db\_xref="GI:4538953"  
/translation="MKIQCEVCEKAEAEVLCCSDAEVLCRPFQIKVHEANKLIFQRHHR  
VALORDASATYAGAPLADICQERKGFLELEBRAMLCNDCAIHTCNHOFLLS  
GVQVSDSLTENSECSTSFSSSEFYQIQSKVLSNYSSEETEGNSGEIYHKNPSVIL  
SP"  
complement(114817..114877)  
/gene="F24G24.40"  
/number=1  
complement(114888..115088)  
/gene="F24G24.40"  
/number=2  
16099..16686  
/gene="F24G24.50"  
/number=1  
16099..16686  
/gene="F24G24.50"  
16099..16686  
/gene="F24G24.50"  
/note="contains EST gb:AA597970"  
/codon\_start=1  
/product="heat shock protein 22.0"  
/protein\_id="CAB39778.1"  
/db\_xref="PID:e1420113"  
/db\_xref="PID:g4538954"  
/db\_xref="GI:4538954"  
/translation="MMKHLISFFIGALLIGNITSEGLSSALETPPGSLISDLMLD  
RFPDFKILRIPLGLERDITVALSPARVDKETAEGHEIMLDIPGLKDKVKEVEE  
NGVLRSGERKEEEKGDQWRRVRSYGFQFKEPLDNDMSVAKKLENGVLTIN  
LTKLSPKVGPRVAVNTAAEDOTAKISSSSKEL"  
complement(117048..117182)  
/gene="F24G24.60"  
/number=1  
17048..18332  
/gene="F24G24.60"  
complement(join(17048..17182,17267..17383,17462..17596,  
17671..18081,18156..18332))  
/gene="F24G24.60"  
/EC\_number="2.7.1.4"  
/note="strong similarity to fructokinase, Lycopersicon  
esculentum, gb:U62329; Contains pfkb family of

Query Match 82.9%; Score 17.4; DB 7; Length 99856;  
Best Local Similarity 94.7%; Pred. No. 2.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 gaaaggaagcctaaagt 21  
|||||  
Db 41288 GAAAGGAAATGCTAAAGT 41270

RESULT 13  
SCHISTRE 12388 bp DNA PLN 29-APR-1998  
LOCUS S.cerevisiae genes for histone H2A and H2B, trehalase, and  
DEFINITION hexaprenyl pyrophosphate synthetase.  
ACCESSION Z26494  
NID 9403311  
VERSION 226494.1 GI:403311  
KEYWORDS centromere 2; hexaprenyl pyrophosphate synthetase; histone H2A;  
histone H2B; trehalase.  
SOURCE baker's yeast.  
ORGANISM Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
Saccharomycetaceae; Saccharomycetes.  
REFERENCE 1 (bases 1 to 12388)  
Wolfe,K.H. and Johan,A.J.E.  
TITLE Sequence around the centromere of Saccharomyces cerevisiae  
JOURNAL chromosome 11: similarity of CEN2 to CEN4  
Yeast 10, 41-46 (1994)

REFERENCE 2 (bases 1 to 12388)  
AUTHORS Wolfe,K.H.  
TITLE Direct Submission  
JOURNAL Submitted (22-SEP-1993) Kenneth H. Wolfe, Department of Genetics,  
University of Dublin, Trinity College, Lincoln Place Gate, Dublin,  
IRL-2, Ireland  
FEATURES  
source  
1..12388  
location/Qualifiers  
/organism="Saccharomyces cerevisiae"  
/strain="S288C"  
/db\_xref="taxon:4932"  
/chromosome="2"  
1..3843  
/clone="cosmid alpha 308.2"  
/gene="YBL0101"  
<1..3843  
/gene="YBL0101"  
/codon\_start=1  
/product="unknown"  
/protein\_id="CA81266.1"  
/db\_xref="PID:g403312"  
/db\_xref="GI:403312"  
/db\_xref="SWISS-PROT:P35194"  
/translation="DHVRSRLISLSILKGLKLEQNDTORILKILVNYNCS  
MSDIEELIYTSISLFTFDERNDVSTELFELGKRPPELISIKLVADLANSYSSR  
MHEYDFRILSTFGKLEIDGKYSSELEMLPLFTFLHFTNNKEELALFNASHAIWK  
FIDPINKRPNLNEASKSISMLKDILPNIRIGRDSLEEQSYVSVLSYVNTKYF  
TDPEDNALILYNGDEADPFTNNHIOHLHROQAIKRLGHAHQKDNSISYLIPI  
EHYVSDDERNNINGETQTAIGLAQHSWNOYKALRYISMLTKPVMQAVOL  
IYOLSVPLRRLIVRGASAKSLTSPSNLDEPSNFYQELIYPLSLKLGTRDET  
IIEMYPALAEVLNIVGLINDDITNFPSTLITNCOYLRSSEELDAVAVTLGKSI  
ILGAEYLVYFKELMATLKRGSQIHVYHVLILKSMHGVLSHSDLPSSMIVKI  
MENIFGAGESEKSEMYHTPKETIKSKSYDAEELIASNLSLEPGTILSPVALLMV  
RINLRNOKLSIELRLGLNNSDSESSILKFOQLFQESMSESPDIPKKYKD  
QVDEKEDFVLNLESKSYTINSNLSLNTLQFADLDFKNVITTRHKSFTVSHLGEF  
IPLRLDLSLENGVYISTLRILITLRIDFSDSESEIFENCARKYLNITKVSPTS  
ELCOMGLKFLSAFIRHTDSTLKDTALSYVGRVLPDLEPSRQGLAFNFKALVSKHI  
MLPELYADIATREIMVNTSKERIDVRSVYQFLMEYDOSGRLEKOPFVNDIO  
YTPESGOSVMELEINLITKANPALLSKSSSFFLAVNVSFNDAPRCHEMSVILS  
TMPLKLENKLEIYEKTIAMKQVNDASLNLGLRIYKYLKISIGEHEIIEDELA  
KRIRYIISDTSVSEHOMDLVYSALNTFSYMEATESVYHGRKDIWDGIIITCLIPH  
SWRQSANLVHQLINANKRLEISLNTLEIOTATRLHQLGAPSIPENIANYSIKTL  
VNISILMKQERPEFIMVSKOTGEDIKYTAIDYMYTRLIGGIIIRSDHRMDSMKA  
CIOLALLVOVLEDEVIAGEKILPLPYGLTYYSRAVDEQEEELRTLSNCKTL  
EDKLOVSDFTKITVAVKQVLESREKRSKRALLVANAPQISADKKIKHARSREKX  
HEEDENKTYRRNKKRKA"  
complement(4120..4518)  
/gene="H2A2 (YBL0103)"  
/db\_xref="SGD:S0000099"  
complement(4120..4518)  
/gene="H2A2 (YBL0103)"  
/codon\_start=1  
/db\_xref="SGD:S0000099"  
/product="histone H2A"  
/protein\_id="CA81267.1"  
/db\_xref="PID:g403313"  
/db\_xref="GI:403313"  
/db\_xref="SWISS-PROT:P04912"  
/translation="MSGGKGGKAGSAAKASOSASAKAGLTFPVGRVHRLIRGNVYAO  
IGGAPVYLTVAVLEYLAETILELAGNARNARNTKRTIIPRLQAIIRNDELINLIGNV  
TIAOGGVLPNINHONLIPKRSATKASQEL"  
5218..5613  
/gene="H2B2 (YBL0104)"  
/db\_xref="SGD:S0000098"  
5218..5613  
/gene="H2B2 (YBL0104)"  
/codon\_start=1  
/db\_xref="SGD:S0000098"  
/product="histone H2B"  
/protein\_id="CA81268.1"  
/db\_xref="PID:g403314"  
/db\_xref="GI:403314"  
/db\_xref="SWISS-PROT:P02294"

```

gene      /translation="MSSAAEKRPASAPAEKKPAAKKTSTVDGKRKRKREYTSY
          IRLKVLQHTDGTISQKSMISLNSFVNDIFERLATRSKLAAYNKRKSTISAREIQYAV
          RLPLGELAKHASECTRAVTKYSSSTOA"
          complement(5878.6192)
          /gene="YBL0105"
          /codon_start=1
          /product="unknown"
          /protein_id="CAA81269.1"
          /db_xref="GI:403315"
          /db_xref="PID:9403315"
          /db_xref="SWISS-PROT:P35195"
          /translation="MPKICLADVCAWPICTDSASISDFVALIEKKIRRESPLTSLHS
          ACTTGPMDDVGLIGEIHGYHEGKGYVAVHDINRGVTRDKHQADRIDVLAKEI
          SQ"
          complement(7666.10008)
          /gene="YBR0106"
          /note="homolog of NTH1 (neutral trehalase) gene on
          chromosome IV (accession number X65925)"
          /codon_start=1
          /product="trehalase"
          /protein_id="CAA81270.1"
          /db_xref="PID:9403316"
          /db_xref="GI:403316"
          /db_xref="SWISS-PROT:P35172"
          /translation="WVDFLEKVTIENPSPSGNDEGNIKPLSSGSEQRPLKEGQOOG
          RHHRLSSWHEEYDFPESNAEYVYPTDPRKOSKILRLRRTMSVFNKSPDEKGM
          KDYTLRSGSDEDSFLSSGSRREYIDNDLADLDELASDPTKNOITREDGPIVI
          KVTANSNGKHNVRGTYSLSLLELITAKSFGKHQITLDEARINENPVDSLSLI
          TIOFWTSLTRVDLYNIAETARDSKIDTPEAKNPRIYVPCDEQYEFYIQASQNPIS
          LKLEVEYLPKIDIAEYKSLNDTPGLLALAMEHVPSTGERSLVGPVAVPGGRNE
          LYGWDSYLMALGLIESNKYDVARGVMEHFEIDHYSKILNARSYLCSPGFLTD
          MALVFEKIGKKNPNAIOLKRAPRAIKYEYVWSSPRLSLISGCHSDGIGI
          PEPTEPDHDTLILPYAEKYNVLEKRIYKNGMKKEPRLDLPFLHDAVRVRSQDIT
          TYREVCAYLATIDINSLLIKYEKIDIAFYIKYFGHEIKDENQVTDSEHNEIAE
          LRRTIRNKYMWDESGFEFFYNTKLCRFYSATSTGSLVAGPVAVPGGRNE
          POLEMLGVLACTEKSRRGPISIDRPIROMDYPPGMAHQIILAKKGSAGYQOAVRL
          AYRLMYITKSPFYDNGVMEKYDYTRGTGTPHVDVAYGQADFRGVATEGFGWNT
          SYLLGKYNMNHARALACSPPLPFPNSLAKPEKRLYYI"
          complement(10433.11293)
          /gene="YBR0107"
          /complement(10433.11293)
          /gene="YBR0107"
          /note="homolog of unidentified ORFs in corynebacterium
          (107603) and Streptomyces (M29297)"
          /codon_start=1
          /product="unknown"
          /protein_id="CAA81271.1"
          /db_xref="PID:9403317"
          /db_xref="GI:403317"
          /db_xref="SWISS-PROT:P35196"
          /translation="NETDSGIRPCHSFYLAKTNTKIFSRITLANSVCVRRHGFINDGKR
          FARKKEDYKEGHEAGFVSMRSLILECEIGVDPAVFAVSTENFRSSSEVSLMTL
          AREIRQITERGEGLACKYGRKRIIDSLDLSLDEVDVAVEIKNNKRALNLNCF
          PYTGREILHAMKETIYOHKGAIDESTESHLYAGVPPDLLLIRTSQVSLSTPL
          IMOASSKGVAILDLCLMPFEGPIRMAMWILKRSFKHSFLNKRYRLDEGQYDETDNGD
          PIDLKEKKLN"
          complement(11534.12388)
          /gene="COO1 (YBR0109)"
          /db_xref="SGD:S0000207"
          /db_xref="PID:9403318"
          /gene="COO1 (YBR0109)"
          /note="overlap with accession J05547"
          /codon_start=1
          /db_xref="SGD:S0000207"
          /product="hexaprenyl pyrophosphate synthetase"
          /protein_id="CAA81272.1"

```

```

          /db_xref="PID:9403318"
          /db_xref="GI:403318"
          /db_xref="SWISS-PROT:P18900"
          /translation="MROBSGAHHIKLISRRRCRSPFAVALNASKRLVTPKILMN
          PISLVSKENKNTLAKNTVALIGSGHPLKNTVYTSIFTEPEKKARPLVLLLSRLSEIP
          WTERNHKLIDKSDVPEPDTYSKPSQNLQORPASSTSPHLILHGLKPLKRGPEPL
          PETEDFKRGITLIPKORLAEIVEMITHATLDDVDYIDSHDRGPRSGNAFTNKAV
          LAGDFILGRATVYSIRLHNPEVVELMSNIAVLVEGEFQIKNTSIDADITIEHGK
          LLPVPSKRL"
BASE COUNT      3949 a      2392 c      2162 g      3885 t
ORIGIN
Query Match      82.9%; Score 17.4; DB 7; Length 12388;
Best Local Similarity 94.7%; Pred. No. 3.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      1 atgaaaggaagctaaaa 19
          |||||||||||||||
Db      42 ATTTAAAGAAAGCTAAA 60
RESULT 14
SCYBL004W      14999 bp      DNA      PLN      11-MAR-1998
LOCUS
DEFINITION
S.cerevisiae chromosome II reading frame ORF YBL004W.
ACCESSION
NID      5438857
VERSION
KEYWORDS
SOURCE
ORGANISM
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE
1 (bases 1 to 2817)
Delaveau,T., Jacq,C. and Perea,J.
Unpublished
REFERENCE
2 (bases 2781 to 13436)
Rieger,M.
Unpublished
REFERENCE
3 (bases 10878 to 14999)
Lohan,A.J.E. and Wolfe,K.H.
Unpublished
REFERENCE
4 (bases 1 to 14999)
MIPS.
Direct Submision
Submitted (30-AUG-1994) Data collected by MIPS on behalf of the
European yeast chromosome II sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
5 (bases 1 to 14999)
Feildmann,H., Aigle,M., Aljinovic,G., Andre,B., Baclet,M.C.,
Barthe,C., Baur,A., Becam,A.M., Bileau,N., Boles,E., Brandt,T.,
Brendel,M., Brueckner,M., Bussearea,F., Christiansen,C.,
Contreras,R., Crouzet,M., Czapluch,C., Demolis,N., Delaveau,T.,
Dolignon,F., Domdey,H., Dueterhus,S., Dubois,E., Dujon,B., El
Bakkouri,M., Ertlan,K.D., Feuermann,M., Fiers,W., Fobo,G.M.,
Fritz,C., Gassenhuber,H., Giansdorff,N., Gottlieb,A., Grivall,L.A.,
de Haan,M., Hein,C., Herbert,C.J., Hollenberg,C.P., Holmstrom,K.,
Jacq,C., Jacquet,M., Jauniaux,D.C., Jonnaux,J.B., Kallese,T.,
Klesau,P., Kirchuth,L., Koetter,P., Korol,S., Liebi,S., Loghe,M.,
Lohan,A.J.E., Louis,E.J., Li,Z.Y., Maat,M.J., Mallet,L.,
Mannhaupt,G., Messenguy,F., Miosga,T., Molmans,F., Mueller,S.,
Nast,F., Obermaier,B., Perea,J., Pierard,A., Pivaandl,E.,
Pohl,F.M., Pohl,T.M., Potier,S., Proft,M., Purnelle,B., Ramezani
Rad,M., Rieger,M., Rose,M., Schaaff-Gerstenschlaeger,I.,
Scherens,B., Schwarlose,C., Stala,J., Stoulski,P.P.,
Smits,P.H.M., Souclet,J.L., Steensma,H.Y., Stucka,R.,
Urrutazazu,A., van der Aart,Q.J., van Dyck,L., Vassarotti,A.,
Vetter,I., Vlerendeels,F., Visser,S., Wagner,G., de Wergifosse,P.,
Wolfe,K.H., Zagulski,M., Zimmermann,F.K., Mewes,H.W. and Kleine,K.
Complete DNA sequence of yeast chromosome II
EMBO J. 13 (24), 5795-5809 (1994)

```

MEDLINE 95112788  
COMMENT On Mar 9, 1999 this sequence version replaced gi:535979.  
FEATURES  
location/Qualifiers  
1. 14999  
/organism="Saccharomyces cerevisiae"  
/db\_xref="taxon:4932"  
/chromosome="II"  
539..639  
/note="delta-remnant"  
635..639  
repeat\_unit  
/note="5 nt direct repeat GGAAT; Tyl target sequence"  
640..973  
/note="5' delta Tyl"  
933..6201  
/gene="TY1B"  
933..2255  
/gene="TY1A"  
933..2255  
CDS  
/codon\_start=1  
/protein\_id="CA84820.1"  
/db\_xref="PID:g535980"  
/db\_xref="GI:535980"  
/db\_xref="SPTREMBL:Q12266"  
/translation="MESQQLSQHSPIFHGSACASVTSKEVQTODPLDISAKTECE  
KVSTQANSQOFTPPSSAVPENHHASPOAOPVLPONGPYPOORMMNTQOANSQGP  
VYGHPSLMPYPPYOMSPYAPPGASQOFTQYQYVGHNTPPESGNSPPDSASAKS  
NMSTNOHVRRPPLITSPNDLNMWKIYIKFLQNSNGDIIPYATRAVQMDDELTA  
FLCHTQFLFASQFLPFWKDLISVDYTDIMKILSKSNKQSDTOEADNITLATAH  
YNGSTPADAEAEVNTILRLNNGGIFLNKAVACOFIMRGLSGEYKFLPYARHCIM  
TVADLFSDHSMYEEOQESKRKSTYRSPDEKDSRTYNTPKSITRNSQKPN  
SOSTRAHNVSTFNNSPGPNDDIRGSTTPEPQLKTMHDLGDELSTGYNTNRS  
DDELPGHLLDSGASRTILRSAHITHSASSNPDIINVDAQKRNIPINALGDLQHFED  
NTKSIYVLTNPNIAYDLISNELAAVDITACFTKNVLEKSDGTVLAPIVKYGDFYV  
SKKYLPSNISVPTINNVHSTSESTRKYPPYIHMMLAHNAQITRSLKNNITTYFE  
SDVMSAIDQCDCLIGKSTRHIGSKRLKONSPEPOYLATDIFGVNHLPRS  
ASYSIEPTDETKFRWYVPLHRRRDSILDPTTILAFIKNOFASVLYOANRSGE  
YNTRLKLEKNGITPCYTTTADSRAGVAERLRLDCCRQLOCSGLPNHIMS  
AIERSTYVRNSLASPKSKAKRQAGLAGDITSLPPGQVYIYNDHNPNSKIPKRI  
PGVALHPSRNSYGIITLPSLKTVDVTNNYILOGKSRLOQFNYDALITDEDNRLT  
ASYSFASNEIOQSNDLNIESDPOSDIYPEOPRNYSKAVSPDTPSPETHE  
DSKRSKTNINAPREVDNISNLSLPSKRSRSPQSIDESTSGCMGKADYVLA  
MSQNTHESSYASKDEBRHSDYSNDENTNTWPISTGGTNNKTPQNSQSEOTER  
TIHRSPIDTSSSSNSLHVPYIKTSDTCCKENTESSIADLPPLPPEPPYELSD  
SFKLPPINSRQNTSLGGIGDSNAYTTINSKRSLEDNETEIVASDNTKMRNL  
EPPRSKRRIHLIAVKAIVKSIKIRITRLIDEATLYNKDIKEKRYIEAHRKQNL  
KMTNWDKYRDEIDPKRVINSMFENRRKDDTHKARFARAGDIQHPDYDQMS  
NTVHYALMTLSLALDNVNIITOLDISAVLYADIKELIYRPPHIGMNDKILRK  
KSLYGIOSGANMVEITIKSYLIRKOGMEVVGSCVFNQVTCIFVDMVLESKRL  
NSKRITDKLMQYDTRKINIGESDEDEIYDILDEIKYQKATMKIGMSLLEKLP  
KLVNPLNKGKRLSPGQGLYIDQLELEDEDYKMKVHEMQQLIGLAVSGKFER  
DLVYINTLAQHILFSPKQYLDMTYELIOFIMNRDQQLIMHKSQPKVPTKLVVID  
ASYGNOPYKSOIGNIYILNGKVIIGKSTKSLSTTEATIEAISVPLNNLSY  
LIOELDKKPIKGLTDSKSTIIISINNEKFPNRFPGTKAMLRDVEGSHLVQY  
IETKKNIAADVMTKRLPIKTFKLTKNKH"

CDS

/codon\_start=1  
/protein\_id="CA84820.1"  
/db\_xref="PID:g535981"  
/db\_xref="GI:535981"  
/db\_xref="SPTREMBL:Q12490"  
/translation="MESQQLSQHSPIFHGSACASVTSKEVQTODPLDISAKTECE  
KVSTQANSQOFTPPSSAVPENHHASPOAOPVLPONGPYPOORMMNTQOANSQGP  
VYGHPSLMPYPPYOMSPYAPPGASQOFTQYQYVGHNTPPESGNSPPDSASAKS  
NMSTNOHVRRPPLITSPNDLNMWKIYIKFLQNSNGDIIPYATRAVQMDDELTA  
FLCHTQFLFASQFLPFWKDLISVDYTDIMKILSKSNKQSDTOEADNITLATAH  
YNGSTPADAEAEVNTILRLNNGGIFLNKAVACOFIMRGLSGEYKFLPYARHCIM  
TVADLFSDHSMYEEOQESKRKSTYRSPDEKDSRTYNTPKSITRNSQKPN  
SOSTRAHNVSTFNNSPGPNDDIRGSTTPEPQLKTMHDLGDELSTGYNTNRS  
DDELPGHLLDSGASRTILRSAHITHSASSNPDIINVDAQKRNIPINALGDLQHFED  
NTKSIYVLTNPNIAYDLISNELAAVDITACFTKNVLEKSDGTVLAPIVKYGDFYV  
SKKYLPSNISVPTINNVHSTSESTRKYPPYIHMMLAHNAQITRSLKNNITTYFE  
SDVMSAIDQCDCLIGKSTRHIGSKRLKONSPEPOYLATDIFGVNHLPRS  
ASYSIEPTDETKFRWYVPLHRRRDSILDPTTILAFIKNOFASVLYOANRSGE  
YNTRLKLEKNGITPCYTTTADSRAGVAERLRLDCCRQLOCSGLPNHIMS  
AIERSTYVRNSLASPKSKAKRQAGLAGDITSLPPGQVYIYNDHNPNSKIPKRI  
PGVALHPSRNSYGIITLPSLKTVDVTNNYILOGKSRLOQFNYDALITDEDNRLT  
ASYSFASNEIOQSNDLNIESDPOSDIYPEOPRNYSKAVSPDTPSPETHE  
DSKRSKTNINAPREVDNISNLSLPSKRSRSPQSIDESTSGCMGKADYVLA  
MSQNTHESSYASKDEBRHSDYSNDENTNTWPISTGGTNNKTPQNSQSEOTER  
TIHRSPIDTSSSSNSLHVPYIKTSDTCCKENTESSIADLPPLPPEPPYELSD  
SFKLPPINSRQNTSLGGIGDSNAYTTINSKRSLEDNETEIVASDNTKMRNL  
EPPRSKRRIHLIAVKAIVKSIKIRITRLIDEATLYNKDIKEKRYIEAHRKQNL  
KMTNWDKYRDEIDPKRVINSMFENRRKDDTHKARFARAGDIQHPDYDQMS  
NTVHYALMTLSLALDNVNIITOLDISAVLYADIKELIYRPPHIGMNDKILRK  
KSLYGIOSGANMVEITIKSYLIRKOGMEVVGSCVFNQVTCIFVDMVLESKRL  
NSKRITDKLMQYDTRKINIGESDEDEIYDILDEIKYQKATMKIGMSLLEKLP  
KLVNPLNKGKRLSPGQGLYIDQLELEDEDYKMKVHEMQQLIGLAVSGKFER  
DLVYINTLAQHILFSPKQYLDMTYELIOFIMNRDQQLIMHKSQPKVPTKLVVID  
ASYGNOPYKSOIGNIYILNGKVIIGKSTKSLSTTEATIEAISVPLNNLSY  
LIOELDKKPIKGLTDSKSTIIISINNEKFPNRFPGTKAMLRDVEGSHLVQY  
IETKKNIAADVMTKRLPIKTFKLTKNKH"

CDS

/codon\_start=1  
/protein\_id="CA84820.1"  
/db\_xref="PID:g535981"  
/db\_xref="GI:535981"  
/db\_xref="SPTREMBL:Q12490"  
/translation="MESQQLSQHSPIFHGSACASVTSKEVQTODPLDISAKTECE  
KVSTQANSQOFTPPSSAVPENHHASPOAOPVLPONGPYPOORMMNTQOANSQGP  
VYGHPSLMPYPPYOMSPYAPPGASQOFTQYQYVGHNTPPESGNSPPDSASAKS  
NMSTNOHVRRPPLITSPNDLNMWKIYIKFLQNSNGDIIPYATRAVQMDDELTA  
FLCHTQFLFASQFLPFWKDLISVDYTDIMKILSKSNKQSDTOEADNITLATAH  
YNGSTPADAEAEVNTILRLNNGGIFLNKAVACOFIMRGLSGEYKFLPYARHCIM  
TVADLFSDHSMYEEOQESKRKSTYRSPDEKDSRTYNTPKSITRNSQKPN  
SOSTRAHNVSTFNNSPGPNDDIRGSTTPEPQLKTMHDLGDELSTGYNTNRS  
DDELPGHLLDSGASRTILRSAHITHSASSNPDIINVDAQKRNIPINALGDLQHFED  
NTKSIYVLTNPNIAYDLISNELAAVDITACFTKNVLEKSDGTVLAPIVKYGDFYV  
SKKYLPSNISVPTINNVHSTSESTRKYPPYIHMMLAHNAQITRSLKNNITTYFE  
SDVMSAIDQCDCLIGKSTRHIGSKRLKONSPEPOYLATDIFGVNHLPRS  
ASYSIEPTDETKFRWYVPLHRRRDSILDPTTILAFIKNOFASVLYOANRSGE  
YNTRLKLEKNGITPCYTTTADSRAGVAERLRLDCCRQLOCSGLPNHIMS  
AIERSTYVRNSLASPKSKAKRQAGLAGDITSLPPGQVYIYNDHNPNSKIPKRI  
PGVALHPSRNSYGIITLPSLKTVDVTNNYILOGKSRLOQFNYDALITDEDNRLT  
ASYSFASNEIOQSNDLNIESDPOSDIYPEOPRNYSKAVSPDTPSPETHE  
DSKRSKTNINAPREVDNISNLSLPSKRSRSPQSIDESTSGCMGKADYVLA  
MSQNTHESSYASKDEBRHSDYSNDENTNTWPISTGGTNNKTPQNSQSEOTER  
TIHRSPIDTSSSSNSLHVPYIKTSDTCCKENTESSIADLPPLPPEPPYELSD  
SFKLPPINSRQNTSLGGIGDSNAYTTINSKRSLEDNETEIVASDNTKMRNL  
EPPRSKRRIHLIAVKAIVKSIKIRITRLIDEATLYNKDIKEKRYIEAHRKQNL  
KMTNWDKYRDEIDPKRVINSMFENRRKDDTHKARFARAGDIQHPDYDQMS  
NTVHYALMTLSLALDNVNIITOLDISAVLYADIKELIYRPPHIGMNDKILRK  
KSLYGIOSGANMVEITIKSYLIRKOGMEVVGSCVFNQVTCIFVDMVLESKRL  
NSKRITDKLMQYDTRKINIGESDEDEIYDILDEIKYQKATMKIGMSLLEKLP  
KLVNPLNKGKRLSPGQGLYIDQLELEDEDYKMKVHEMQQLIGLAVSGKFER  
DLVYINTLAQHILFSPKQYLDMTYELIOFIMNRDQQLIMHKSQPKVPTKLVVID  
ASYGNOPYKSOIGNIYILNGKVIIGKSTKSLSTTEATIEAISVPLNNLSY  
LIOELDKKPIKGLTDSKSTIIISINNEKFPNRFPGTKAMLRDVEGSHLVQY  
IETKKNIAADVMTKRLPIKTFKLTKNKH"

repeat\_unit

/gene="TY1A"  
/note="5 nt direct repeat TGGTA"

repeat\_unit

/gene="TY1A"  
/note="5 nt direct repeat TGGTA"

misc\_feature

2235..2241  
/gene="TY1A"  
/note="Tyl frameshift CTAGGC"

repeat\_unit

6219..6223  
/note="5 nt direct repeat TGGTA"

LTR

6224..6555  
/note="3' delta Tyl"

repeat\_unit

6556..6560  
/note="5 nt direct repeat GGAAT; Tyl target sequence"

tRNA

6678..6759  
/gene="ts(UC)B - systematic name"

gene

6678..6759  
/gene="ts(UC)B - systematic name"

CDS

7239..14720  
/note="orf YEL004"

CDS

/codon\_start=1  
/protein\_id="CA84821.1"  
/db\_xref="PID:g1263935"  
/db\_xref="GI:4388558"  
/translation="MAKQRTTSSKRYRYSFKARIDDLKIEPARNLKRYHYVES  
SHFLASFDQKKEINLSAKTERFAETHEVDIPLQILYHDKITFNSLVSFINFHDES  
LQPLDLAQFCHDLPDFLKEFEAKITLINDLAIEFESSNVEWEGFNCLAYTEK  
YLSKFLVKKLYVCDLIPILSHKSELSRFSKESALFVRKCPVSNLEFEVRSYK  
LEEDDQRTNLYELLIFTESMSTOETLSKAKAIMSVLHVALTSKSPERSVLS  
DIMNISKASIESLIPYEVWQDNDSLANDRIDRLIKVLTITVFSGGRIPWNI  
KITILIERKMSOENASISODVATPLPALTINSVKTITLPHQKLPYALNIDC  
ELFEFQALKSTERYFSNGLEFLQFLKKNQSGKTALEFLEVDKPELQKRE  
VNPPEFTLISROFETVATINDSNDLEIYWRALIKRYKLONTIILPLEIFSTF  
ASPDNRTKMWGTLTIYKREDASNNELITLDYENYKESLNLKRWMLKVSILH  
PSESLGMSHYPSLLSLDNTFMDPKIRYETLEMTLMILOQMOVDLSGCM  
IEEIPILQIARBDITIRKNVGAEGKRTDKVSEFPLKYLEGILLTVRRSPVQGV  
DTIPNYTDEALVMTKYLSFTLDPENONLDYQGLLEDGAKVYIMDSVYRLPTI  
DYSHIMSKSTONISTITIRKGNITPILIRKQALKMLISTKYVAENHEDVIAF  
EYVNDERKTKDEEDMENERITGTSWTVRNVFLKTLSEFKNTQVSALEHDLIAV  
LIGSRNDVOKIALDALAYKNPTLKNRYKNDNLKLDLDFLKEIDITFLENGSQIK  
AEEKRYVMPYVILIFFGAAOVPTSGSKSRKJAVSYVNFEPKPIYINPLSIASRL  
DYVFFGNSHQIINSSKATLTITIRMTGPNVYVSTLSTVIRTNPLTNSVLOPLIYI  
AMYYIYLDNESTREYHTRKAKSLRQOGLKCLSSVVEFGNTFDWSTMEDYAVYK  
PRISHSDEMLQOPSSLRFLFLWAINPSLYOPLYDERATRALDITISNOVKAVY  
IGPIEABOSIINNPNVDHYVLDVLTICTCLKILPSYVTKLSDNSISTELNLLVS  
ITMGGIYODHVSRISSILSLIKKRLKQENDOKIKLIKILVFNYSMSWDIE  
EKTITTSLEKTEYDERNLNLELIEIGRKYPELESISKVADNSYSSSMHDE  
FPLIISFEGKLEIDGKYSSELMLELFTFLFHEINNKELALRTVASHAKMFTDI  
NEPNINAKSKSISMLKIDILPIRIRIGLDSLEVOGSEVYSVLSYVKNYKTYTDED  
MALLLNGBEADFTVNVNIQLRKQRAIKRGHAHOLKDSISHYILPMIEHYV  
SDERYRNIGNETQIAIGLAQMSNQQKALRLRYISMLKTPNOMKQAVOLIVOLS  
VPJRETLRIYRDAESKRLTSKSPSLDESPNSIRKELPYTLKILIGTDDDETIIEM  
PIAEALVNIYLGATNDIINFLEPSILTNICOVYLRKSEELRDVARTLGKISTIGAE  
YIVFVIRKEMALITKRGSOIHVLSYVAYGILSMHGVALKSGDITSSMILKIMEF  
GFRGERKDSNHYTKYKELIKSNKSYAGELIKASNISITFEGTILSYVKALLMYRILR  
NORKLELRLRYTLGLNHSNDSSESILFCHQDFOESENSPQIPKRVKQVOEK  
EDFLELVKESKTYINSNLLNLSLQKALDLIRNATYRHRFLVSHLGEIPLR  
DSLLENGEYSTILRLITLILSLDSDBSSEELFNKCAKRVLIIVKVSSTSEELCM  
GLFEISAFIHTSTKDALSVLYRVOYFIMDESPROGALFAFELKALSHIMEPL  
YDIAADTRETIMNHGKEIRDOVRSYVOTFMDKQSRLEOKLEOFAMVDNIOYPER  
GROSVNELIULITKANPALSLSSFFALVYNSFNDAAPRCMAASYLSTMPK  
LEKDLIEYKRYIAAMUKOVNNSFNLNGIRYKVLKISGEHTIIELEDLAKRLRY  
ILSDTSVSGSHOMDLVYSALNTSSTMEATNERYKVLKISGEHTIIELEDLAKRLRY  
SAANILHOLANDKLEISLTNEIOTIATRIHLOGASIPENLANSIKTLVANSI  
LMEORPPEFMDVSKOTGEKLYTALIDVWTRIGGILISDERHMSFMSKACIOL  
ALIVOUDEYVABEKIILPYGYLETTYSSAVDEDELETPTLSNECLKILEDKO  
VSPFTKITVAKQTVYLERKRRRSKAIILAVNAPQISADKILKRAHRSREKRRHEDE  
NGYORRNRKRRR"

BASE COUNT 5064 a 2807 c 2620 g 4508 t  
ORIGIN

Query Match 82.98; Score 17.4; DB 7; Length 14999;  
Best Local Similarity 94.7%; Pred. No. 3.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atgaaaaggaagctaaaa 19  
 Db 10919 ATTAAGCAAGCTAATA 10937

RESULT 15  
 T9A4 83922 bp DNA PLN 06-OCT-1998  
 LOCUS Arabidopsis thaliana BAC T9A4.  
 DEFINITION AF096373  
 MID 93695400  
 VERSION AF096373.1 GI:3695400  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE 1 (bases 1 to 83922)  
 AUTHORS Washington University Genome Sequencing Center.  
 TITLE The A. thaliana Genome Sequencing Project.  
 JOURNAL Unpublished (1997)  
 REFERENCE 2 (bases 1 to 83922)  
 AUTHORS Zidanic, M., McQuerry, Y., and Smith, A.  
 TITLE The sequence of A. thaliana T9A4  
 JOURNAL Unpublished (1998)  
 REFERENCE 3 (bases 1 to 83922)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-OCT-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University,  
 St. Louis, MO 63108, USA  
 e-mail: twilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by M. Lohli, Cold Spring Harbor Laboratories, and fingerprinted by M. Maier, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

#### NEIGHBORING COSMID INFORMATION:

The 3' clone is T17F16. Actual start of this clone is at base position 1 of T9A4 ; actual end is at 83992 of T9A4.

#### NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFINDER (P. Green and L. Hillier, ms in preparation).

#### FEATURES

source  
 1. 83922  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="T9A4"

gene  
 CDS  
 /chromosome="IV"  
 /map="unknown"  
 486..3161  
 /gene="T9A4.4"  
 join(486..665,756..839,938..1147,1257..1382,1469..1606,1698..1801,1893..1944,2103..2191,2298..2481,2533..2596,2672..2739,2908..3030,3126..3161)  
 /gene="T9A4.4"  
 /note="similar to Isolucyl-tRNA synthetases"

gene  
 CDS  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="Aac62806.1"  
 /db\_xref="PID:g3695406"  
 /db\_xref="GI:3695406"  
 /translation="MTSLTAFKPAFKNLICNGIVLAEDKKMAKRLRYPPELVI  
 DEGVASFPNMLCCDAVRLYLINSYVAEPLRKEEVLGVADVPLPYNARF  
 LVONAKLETGSGVFPVPIDLATIQSANIIDONIHSAVOSLVFVREMDAYLVYV  
 PRLKRLDNTNTIVFRNRRLRGEDDCHTALSTFNVLTSCVAPETPFTE  
 TLVONLRKACKGSESVYCSIPRREGMEGERELSTYRMKKIIDLARINRENNKLP  
 KTLPLKEMIVVHPADFLNDITGVLEELNANSLVPCNDLTKVASLAKPEPESVGRKL  
 GKMGLIVAKREKMSOKDILAFEEAGVITANHLKETDICKVSHAMVSEIALRIYR  
 VFRPPDLKRENEIDSAGDGVYITLIDRADDSLVKAGFAREIYNRIOLKTKKSGLEPT  
 DFEVYFQSDIDESVSKOYVLSQWLTKRKEPPT"  
 complement(5712..9371)  
 /gene="T9A4.5"  
 complement(join(5712..5939,6786..6918,8299..8988,9046..9371))  
 /gene="T9A4.5"  
 /note="similar to potassium transport proteins"

gene  
 CDS  
 /evidence=not\_experimental  
 /protein\_id="Aac62807.1"  
 /db\_xref="PID:g3695407"  
 /db\_xref="GI:3695407"  
 /translation="MDRVVAKIKAKINSQTLKLSLFFLYLYLFFSFLGFLAKTKT  
 PRITSRHPDLEFTSVATTVSSMSVMDVENSQILFLITLMEGCEIFSPFLN  
 YSVFNTSIEDRCDEVEVDYREGILKMDREKACLYSVYLIPRNLGVSLLVY  
 VNFVKARVDLSKESISPLFSEVTVTFPANGCFVPTNEMIIIPRNSGLIMLLIYQ  
 VMGNLTFPCEVILIMGLIKYIKRDEKCYILKNHMKGSHLSRLCYLLVYVQ  
 FLITQLFCAFERMTSESLEGMSSYEKLVGSLQVNSRHTGETIVDSLTSFVILV  
 FILMIGLIVSLSFLTICFLISITERQNQRPIENFNITILEYIYRFGNSAVGN  
 VGTGTGSCERRVDISDGGCKDASYPFAGRMSPMGRFVLIIVFYGRFOFTAKSGRA  
 WILPSSS"  
 complement(26302..26574)  
 /gene="T9A4.6"  
 complement(26302..26574)  
 /gene="T9A4.6"  
 /note="contains similarity to Solanum lycopersicum (tomato) wound-induced protein (GB:X59882)"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="Aac62808.1"  
 /db\_xref="PID:g3695408"  
 /db\_xref="GI:3695408"  
 /translation="MSSTSKRAKTVAVSIGAVEALKDOLGCRNNTILRSVNOHLRNNV  
 RSVSGKRFSSSSVSAVTSSESEKAKAEELRTVMYLSGCPN"  
 28609..28893  
 /gene="T9A4.3"  
 join(28609..28785,28860..29270,29345..29479,29558..29674,29759..29893)  
 /gene="T9A4.3"  
 /note="contains similarity to the pKb family of carbohydrate kinases (Pfam: PF00294, E-1.6e-75)"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="Aac62803.1"  
 /db\_xref="PID:g3695403"  
 /db\_xref="GI:3695403"  
 /translation="MANPPIVYSGEMLDIYPTDTSVGSIAESTGFLKAPGAPANVA  
 CATIKLGRSAFVIGKFGDEFGHMLVNIILKKNVNSGVCFFDNRATATAFVTLKDG  
 EREFMFYNPSADMLKESELNDLIRKAKIFHYGISILISEPCRTAANAAKMTADA  
 GVLISYDPNVRPLPSTEAIEIGKISINMEADIIVKSDDEVSLTRGDAEKDQVLS  
 LMHDRKLLIVTGEKGRYTKKFKGRVAVKAVDTTGACDSVGAFLVLSLGDG

Search completed: August 18, 1999, 17:17:01  
Job time: 3467 sec

gene  
CDS  
STLDEGKIKKALAFANAGAVCTTQKAIPLPPPADAKIMKSKR"  
complement(30255..30842)  
/gene="T9A4.7"  
complement(30255..30842)  
/gene="T9A4.7"  
/note="contains similarity to heat shock hsp20 proteins  
(p1am: PP0011, E=1.2e-46"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAC62802.1"  
/db\_xref="PID:g3695402"  
/db\_xref="GI:3695402"

gene  
CDS  
/translation="MKHLISFFIGALLGNIKTSESSLSALETPPGSLISDMLD  
RPPDPFKILERDPLGLERDYSVALSPARYDMKETAEGHEIMLDIPGLKDEKVEIEE  
NGVLRSGERKEEEKGDQDMRVERISYGRFPROFKLPDNDMESVYKAKLENGVLTIN  
LTKLSEPKYKGRVYNIAAEEDQTAKISSESEKEL"  
31853..32412  
/gene="T9A4.2"  
join(31853..32053,32125..32412)  
/gene="T9A4.2"  
/note="contains similarity to Arabidopsis thaliana  
salt-tolerance protein (GB:X95572) and CONSTANS-like 1  
proteins"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAC62805.1"  
/db\_xref="PID:g3695405"  
/db\_xref="GI:3695405"  
/translation="MKIOCEVGEKAEAVILCCSDPAVLCPPGDIKVEANKLFORHHR  
VALQKDAASATTASGAPLCIDICQERKGYFCLIEDRAMLCNDCEAIHTCNSHORFLS  
GVQVSDQSLTENSECSTSFSEFYQISKVSLSNQISSETEAGNSGETVHKNPVIL  
SP"

gene  
CDS  
complement(34194..35118)  
/gene="T9A4.8"  
complement(join(34194..34922,35026..35118))  
/gene="T9A4.8"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAC62809.1"  
/db\_xref="PID:g3695409"  
/db\_xref="GI:3695409"

gene  
CDS  
/translation="MTSILSAVSKLRGCVNQIENKNRPSGASEDILNQAKMLTQYER  
YKRGFEHWPILKIGERFANDNKTTPAFQEGEDVTSSTSSSTINTESSPSPMANS  
IDLNDSEDAWPSLSRPMGLKAKRQOSEOFQQLQNDKLIKAITKGTSENEI  
QROKIEVARMKEENKILFADLNSISDPSRAVYENRKRILKRAQTNOHEDEGGSQ  
YHGOYRASHYQESLPHGKVOGEPDQGEDKRSPPNNQEDFTQYVNYLSGTGNMP"  
complement(36300..39184)  
/gene="T9A4.9"  
complement(join(36300..36656,37387..37539,38026..38214,  
38746..38838,38999..39184))  
/gene="T9A4.9"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAC62813.1"  
/db\_xref="PID:g3695413"  
/db\_xref="GI:3695413"

/translation="MTIGCALPMTCMIIACYLIGCSLMSHCHGVYKEAKTLKSNEDL  
ETEHKLINKPAFKIVKTINGERGCVDFYKOPGLDHSMMKHFHRTKRTTEGH  
FWENGVCIGITVPIDPRYTKDALLRKMSFSDNSNPSQSKTYKPAASIDDHFAVY  
RTTKGTRISNGASMNITFTPSVGPQFSASRHRQIGNEFTQVGHIDKINGNMILM  
GTSWEVGFPSSRFKESGIVEMGEGEYSPSPNPPGNSHYPRKSPKVDSTYVRLI  
TIVDENYNTDKTVKNTERYSDCYVRDATERFWSHGLIYGGPGCK"

Query Match 82.9%; Score 17.4; DB 8; Length 83922;  
Best Local Similarity 94.7%; Pred. No. 2.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 gaaaaggaagcctaaagt 21  
|||||  
Db 5653 GAAAAGGATGCTAAAGT 5671

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 17:18:23 ; Search time 148.09 Seconds  
(without alignments)  
35.479 Million cell updates/sec

Title: us-09-004-395-3

Perfect score: 1 atgaaaggaagctaaagt 21

Sequence: IDENTITY\_NUC

Scoring table: 311585 segs, 125096042 residues

Searched: N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	21	100.0	110000	1	X20248_01	Continuation (2 of
2	21	100.0	110000	1	X20248_02	Continuation (3 of
3	16.8	80.0	2569	1	V21209_08	Continuation (9 of
4	16.8	80.0	2569	1	X40071	Colon cancer assoc
5	16.4	78.1	1288	1	V84525	Human secreted pro
6	16.2	77.1	14382	1	N81546	Bio A, Bio B and B
7	16.2	77.1	14255	1	Q43526	ALL-1 gene cDNA. D
8	16.2	77.1	14201	1	O53475	11q23 chromosome t
9	16.2	77.1	14201	1	O53477	11q23 chromosome t
10	16.2	77.1	11907	1	O45339	cDNA encoding htrix
11	16.2	77.1	14255	1	O75181	ALL-1 (acute lymph
12	16.2	77.1	14201	1	T16330	MLL 1.5EB fragment
13	16.2	77.1	14201	1	T16332	MLL 1.5EB fragment
14	16.2	77.1	110000	1	T58840_0	mycoplasma genital
15	16.2	77.1	1438	1	T51257	Human Ad4 gene gen
16	16.2	77.1	594	1	T90370	Alga acetyl-CoA ca
17	16.2	77.1	2118	1	T90371	Recombinant molecu
18	16.2	77.1	1851	1	T90372	Recombinant molecu
19	16.2	77.1	2224	1	T90373	Human ATL-1/AF-4 o
20	16.2	77.1	14255	1	V20477	Kojibiose phosphor
21	16.2	77.1	3956	1	V31377	Staphylococcus aur
22	16.2	77.1	2325	1	V31376	Staphylococcus aur
23	16.2	77.1	2108	1	V48280	Enterococcus faeca
24	16.2	77.1	1607	1	V75167	5-HT1F receptor cl
25	16.2	77.1	16870	1	X13035	Human MR7 serotonin
26	16.2	77.1	480	1	X13749	Spodoptera ecdyson
27	15.8	75.2	1730	1	O46580	Thermophilic facto
28	15.8	75.2	1554	1	O72272	DNA encoding a the
29	15.8	75.2	948	1	T45796	L-alpha-glyceropho
30	15.8	75.2	948	1	T45796	PKR2-26 containin
31	15.8	75.2	1012	1	V87288	Microbial L-alpha-
32	15.8	75.2	5040	1	X20272	Microbial L-alpha-
33	15.8	75.2	414	1	X15275	Microbial L-alpha-
34	15.4	73.3	1818	1	N90297	Microbial L-alpha-
35	15.4	73.3	12145	1	T38833	Microbial L-alpha-
36	15.4	73.3	1797	1	V35137	Microbial L-alpha-
37	15.4	73.3	1797	1	V35138	Microbial L-alpha-
38	15.4	73.3	1797	1	V35140	Microbial L-alpha-
39	15.4	73.3	1797	1	V35141	Microbial L-alpha-
40	15.4	73.3	1797	1	V35142	Microbial L-alpha-
41	15.4	73.3	1797	1	V35143	Microbial L-alpha-
42	15.4	73.3	1824	1	V35144	Microbial L-alpha-
43	15.4	73.3	1800	1	V35139	Microbial L-alpha-

## ALIGNMENTS

44 15.4 73.3 1824 1 V35133  
45 15.4 73.3 1797 1 V35134

Microbial L-alpha-  
Microbial L-alpha-

RESULT 1  
X20248\_01/c  
Continuation (2 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide  
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
WP Fragment Name Begin End  
WP X20248\_00 1 110000  
WP X20248\_01 100001 210000  
WP X20248\_02 200001 310000  
WP X20248\_03 300001 410000  
WP X20248\_04 400001 510000  
WP X20248\_05 500001 610000  
WP X20248\_06 600001 710000  
WP X20248\_07 700001 810000  
WP X20248\_08 800001 910000  
WP X20248\_09 900001 910715

Query Match 100.0%; Score 21; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaaaggaagctaaagt 21  
DB 105140 ATGAAAGGAAGCTAAAGT 105120

RESULT 2  
X20248\_02/c  
Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide  
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
WP Fragment Name Begin End  
WP X20248\_00 1 110000  
WP X20248\_01 100001 210000  
WP X20248\_02 200001 310000  
WP X20248\_03 300001 410000  
WP X20248\_04 400001 510000  
WP X20248\_05 500001 610000  
WP X20248\_06 600001 710000  
WP X20248\_07 700001 810000  
WP X20248\_08 800001 910000  
WP X20248\_09 900001 910715

Query Match 100.0%; Score 21; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaaaggaagctaaagt 21  
DB 5140 ATGAAAGGAAGCTAAAGT 5120

RESULT 3  
V21209\_08  
Continuation (9 of 17) of V21209 from base 800001 (Methanococcus jannaschii circular  
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209  
WP Fragment Name Begin End  
WP V21209\_00 1 110000  
WP V21209\_01 100001 210000  
WP V21209\_02 200001 310000  
WP V21209\_03 300001 410000  
WP V21209\_04 400001 510000  
WP V21209\_05 500001 610000  
WP V21209\_06 600001 710000  
WP V21209\_07 700001 810000  
WP V21209\_08 800001 910000

WP V21209\_09 900001 1010000  
WP V21209\_10 1000001 1110000  
WP V21209\_11 1100001 1210000  
WP V21209\_12 1200001 1310000  
WP V21209\_13 1300001 1410000  
WP V21209\_14 1400001 1510000  
WP V21209\_15 1500001 1610000  
WP V21209\_16 1600001 1664976

Query Match 80.0%; Score 16.8; DB 1; Length 110000;  
Best Local Similarity 90.0%; Pred. No. 80;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atgaaaggaagctaaag 20  
|||  
Db 45199 ATTAAGCAATCTAAAG 45218

RESULT 4  
X40071  
ID X40071 standard; DNA; 2569 BP.

AC X40071:  
DE 02-JUL-1999 (first entry)  
KW Colon cancer associated gene.  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer; ss.  
OS Homo sapiens.  
PN MO9804265-A2.  
PD 28-JAN-1999.  
PR 15-JUL-1998; U14679.  
PR 22-JUN-1998; US-102322.  
PR 17-JUL-1997; US-896164.  
PR 10-OCT-1997; US-061599.  
PR 10-OCT-1997; US-061765.  
PR 10-OCT-1997; US-948705.  
PR 11-OCT-1997; GB-021697.  
PA (LUDW-) LUDWIG INST CANCER RES.  
PI Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,  
PI Pfundschuh M, Sahin U, Scanlan MJ, Stockert E,  
PI Tureci O;  
DR WPI: 99-132448/11.  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
PS Claim 67; Page 677-678; 787pp; English.  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
SQ Sequence 2569 BP; 865 A; 418 C; 564 G; 712 T;

Query Match 80.0%; Score 16.8; DB 1; Length 2569;  
Best Local Similarity 90.0%; Pred. No. 71;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgaaggaagctaaag 21  
|||||  
Db 1242 TGAAGCAAGGTAAAGT 1261

RESULT 5  
ID V84525 standard; DNA; 1288 BP.  
AC V84525:  
DE 01-MAR-1999 (first entry)  
KW Human secreted protein gene 115 clone HDTAM95.  
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
OS Homo sapiens.  
PN WO9854963-A2.  
PD 10-DEC-1998.  
DE 04-JUN-1998; U11422.  
PR 18-DEC-1997; US-070923.  
PR 06-JUN-1997; US-048877.  
PR 06-JUN-1997; US-048881.  
PR 06-JUN-1997; US-048884.  
PR 06-JUN-1997; US-048893.  
PR 06-JUN-1997; US-048896.  
PR 06-JUN-1997; US-048899.  
PR 06-JUN-1997; US-048915.  
PR 06-JUN-1997; US-048949.  
PR 06-JUN-1997; US-048964.  
PR 06-JUN-1997; US-048972.  
PR 06-JUN-1997; US-049020.  
PR 06-JUN-1997; US-049375.  
PR 05-SEP-1997; US-057628.  
PR 05-SEP-1997; US-057635.  
PR 05-SEP-1997; US-057644.  
PR 05-SEP-1997; US-057647.  
PR 05-SEP-1997; US-057650.  
PR 05-SEP-1997; US-057661.  
PR 05-SEP-1997; US-057667.  
PR 05-SEP-1997; US-057761.  
PR 05-SEP-1997; US-057764.  
PR 05-SEP-1997; US-057770.  
PR 05-SEP-1997; US-057775.  
PR 05-SEP-1997; US-057778.  
PR 06-JUN-1997; US-048875.  
PR 06-JUN-1997; US-048878.  
PR 06-JUN-1997; US-048882.  
PR 06-JUN-1997; US-048885.  
PR 06-JUN-1997; US-048894.  
PR 06-JUN-1997; US-048897.  
PR 06-JUN-1997; US-048900.  
PR 06-JUN-1997; US-048916.  
PR 06-JUN-1997; US-048962.  
PR 06-JUN-1997; US-048970.  
PR 06-JUN-1997; US-048974.  
PR 06-JUN-1997; US-049373.  
PR 05-SEP-1997; US-057584.  
PR 05-SEP-1997; US-057629.  
PR 05-SEP-1997; US-057642.  
PR 05-SEP-1997; US-057645.  
PR 05-SEP-1997; US-057648.  
PR 05-SEP-1997; US-057651.  
PR 05-SEP-1997; US-057662.  
PR 05-SEP-1997; US-057668.  
PR 05-SEP-1997; US-057762.  
PR 05-SEP-1997; US-057765.  
PR 05-SEP-1997; US-057771.  
PR 05-SEP-1997; US-057776.  
PR 06-JUN-1997; US-048876.  
PR 06-JUN-1997; US-048880.  
PR 06-JUN-1997; US-048883.  
PR 06-JUN-1997; US-048892.  
PR 06-JUN-1997; US-048895.  
PR 06-JUN-1997; US-048898.



PR 06-JUN-1997: US-048901.  
 PR 06-JUN-1997: US-048917.  
 PR 06-JUN-1997: US-048963.  
 PR 06-JUN-1997: US-048971.  
 PR 06-JUN-1997: US-049019.  
 PR 06-JUN-1997: US-049374.  
 PR 05-SEP-1997: US-057627.  
 PR 05-SEP-1997: US-057634.  
 PR 05-SEP-1997: US-057643.  
 PR 05-SEP-1997: US-057646.  
 PR 05-SEP-1997: US-057654.  
 PR 05-SEP-1997: US-057659.  
 PR 05-SEP-1997: US-057666.  
 PR 05-SEP-1997: US-057760.  
 PR 05-SEP-1997: US-057763.  
 PR 05-SEP-1997: US-057769.  
 PR 05-SEP-1997: US-057774.  
 PR 05-SEP-1997: US-057777.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,  
 PI Fan P, Feng P, Ferlie AM, Fischer CL, Florence C,  
 PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DM,  
 PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,  
 PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;  
 DR WPI: 99-059865/05.  
 DR P-PSDB: W88648, W88934.  
 PT New isolated human genes and the secreted polypeptides they encode -  
 PT useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 PS Claim 4; Page 377-378; 772pp; English.  
 CC The invention relates to nucleic acid sequences (W84411 to W84633)  
 CC encoding human secreted proteins (W88534 to W88756). The secreted protein  
 CC gene sequences are deposited with the ATCC under deposit numbers ATCC  
 CC 57979, 57974, 57975, 57976, 57977, 209007, 209008, 209009, 209010,  
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host  
 CC cells comprising recombinant vectors containing the nucleic acid  
 CC sequences are used for the recombinant production of the secreted  
 CC proteins. The polynucleotide and amino acid sequences are useful for are  
 CC useful for preventing, treating or ameliorating medical conditions e.g.  
 CC by protein or gene therapy. Pathological conditions can be also diagnosed  
 CC by determining the amount of the new polypeptides in a sample or by  
 CC determining the presence of mutations in the new polynucleotides.  
 CC Specific uses are described for each of the polynucleotides, based on  
 CC which tissues they are most highly expressed in, and include developing  
 CC products for the diagnosis or treatment of cancer, neurodegenerative  
 CC disorders, developmental abnormalities and foetal deficiencies, blood  
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune  
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,  
 CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,  
 CC restenosis, prostate diseases, obesity, disorders involving osteoclasts  
 CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung  
 CC or thymus, digestive/endocrine disorders, infections and AIDS. The  
 CC polypeptides are also useful for identifying their binding partners.  
 CC The present sequence represents a gene encoding a human secreted protein  
 CC (see descriptor line for gene number and clone identification).  
 SQ Sequence 1288 BP; 356 A; 264 C; 295 G; 370 T;

Query Match 78.18; Score 16.4; DB 1; Length 1288;  
 Best Local Similarity 94.48; Pred. No. 1e+02;  
 Matches 1; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgaaaagaagaactaa 18  
 |||||  
 DB 953 ATGAAAACCAACTAA 970

RESULT 6  
 ID N81546 standard; DNA; 4382 BP.  
 AC N81546;  
 DT 04-DEC-1990 (first entry)  
 DE Bio A, Bio B and Bio D-encoding fragment of B.sphaericus IFO 3525.

KW biotin biosynthesis; fermentative biotin production; plasmid pTG1400;  
 KW bio A; bio B; bio D; seborrheic dermatitis treatment; ss.  
 OS Bacillus sphaericus.  
 FH Key location/Qualifiers  
 FT misc\_rna 1..488  
 FT /\*tag- a /note="upstream non-coding region of bio D ORF"  
 FT /\*tag- b 489..1214  
 FT /\*tag- b /product="dethiobiotin synthetase  
 FT /\*tag- c 491..504  
 FT /\*tag- c /note="Bio D"  
 FT /\*tag- d 510..512  
 FT /\*tag- d /label="possible initiation codon  
 FT /\*tag- e 575..577  
 FT /\*tag- e /label="possible initiation codon  
 FT /\*tag- f 584..586  
 FT /\*tag- f /label="possible initiation codon  
 FT /\*tag- g 1194..1196  
 FT /\*tag- g /label="possible initiation codon  
 FT /\*tag- h 1188..1262  
 FT /\*tag- h /product="7,8-diamino pelargonic aminotransferase  
 FT /\*tag- i 1260..1262  
 FT /\*tag- i /label="possible initiation codon  
 FT /\*tag- j 1431..1433  
 FT /\*tag- j /label="possible initiation codon  
 FT /\*tag- k 1574..1576  
 FT /\*tag- k /label="possible initiation codon  
 FT /\*tag- l 2509..3159  
 FT /\*tag- l /label="possible initiation codon  
 FT /\*tag- m 2530..2532  
 FT /\*tag- m /product="unknown  
 FT /\*tag- n 2542..2544  
 FT /\*tag- n /label="possible initiation codon  
 FT /\*tag- o 2566..2568  
 FT /\*tag- o /label="possible initiation codon  
 FT /\*tag- p 2599..2601  
 FT /\*tag- p /label="possible initiation codon  
 FT /\*tag- q 3139..3148  
 FT /\*tag- q /label="possible initiation codon  
 FT /\*tag- r 3141..4154  
 FT /\*tag- r /product="biotin synthetase  
 FT /\*tag- s 3156..3158  
 FT /\*tag- s /note="Bio B"  
 FT /\*tag- s /label="possible initiation codon  
 FT /\*tag- t 3183..3185  
 FT /\*tag- t /label="possible initiation codon  
 FT /\*tag- u 3261..3263  
 FT /\*tag- u /label="possible initiation codon  
 FT /\*tag- v 4317..4333  
 FT /\*tag- v /label="possible initiation codon  
 FT /\*tag- v /function="possible termination site  
 FT /\*tag- v /note="palindromic regions - 4317-4333

FT and 4348-4364"

PN EP-266240-A.

PD 04-MAY-1988.

PF 28-SEP-1987; 402157.

PR 30-SEP-1986; FR-013603.

PR 18-MAY-1987; FR-006916.

PA (TRAN-) Transgene SA.

PI Gloeckler R, Speck D, Lemoline Y.;

DR WPI; 88-121306/18.

P-PSDB; P81188, P81189, P81190, P81191.

PT New DNA sequences encoding enzymes involved in biotin biosynthesis -

PI isolated from *Bacillus sphaericus*, and transformants useful in

PI fermentative biotin production.

PS Disclosure; 6, 7, 8; 51pp; French.

CC DNA was isolated from *B. sphaericus* IFO 3525, cut with HindIII and

CC the fragments sub-cloned into pBR322. The recombinant plasmids were

CC used to transform *E. coli* bio- mutants. Plasmid pTG1400 was present

CC in one clone (*E. coli* C600; CNCK I-608) which showed complementation

CC for bio A and bio D. The plasmid contains a 4.3kb insert which

CC includes 4 long open reading frames (LORFs) as described in the

CC Features Table.

CC A second clone was isolated which contained DNA encoding bio F and

CC bio C. The insert present in the plasmid of this clone and the 4.3kb

CC insert from pTG1400 were cloned together in pBR322 to produce a

CC plasmid (pTG1440) which complements biotin auxotrophy in a bio delta

CC FCD mutant.

CC Transformed cells are used to produce biotin by culture in a medium

CC containing pimelic acid and/or biotin vitamer. Co-culture of two

CC transformants is preferred where one synthesises the vitamer from

CC pimelic acid and the other converts the vitamer to biotin.

CC See also N81551.

SQ Sequence 4382 BP; 1377 A; 745 C; 952 G; 1308 T;

Query Match 77.1%; Score 16.2; DB 1; Length 4382;

Best Local Similarity 85.7%; Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atgaaaaggaaagctaaagt 21

DB 3318 ATGATTATGCAAGCTAAAGT 3338

RESULT 7

ID Q43526/c

AC Q43526;

DT 08-NOV-1993 (first entry)

DE ALL-1 gene cDNA.

KW Acute lymphoblastic leukemia gene; ALL-1; chromosome 11; treatment;

KW translocation breakpoint mapping; chromosomal abnormality; diagnosis;

KW human; acute lymphocytic; myelomonocytic; monocytic; myelogenous;

KW leukemia; ss.

OS Homo sapiens.

FH key

FT Location/Qualifiers

FT 1..1173

FT cds

FT /\*tag= a

FT /note= "ALL-1 gene"

PN WO9312136-A.

PD 24-JUN-1993.

PF 09-DEC-1992; U10930.

PR 11-DEC-1991; US-805093.

PR 27-MAY-1992; US-888839.

PR 30-OCT-1992; US-971094.

PA (UYE-) UNIT JEFFERSON THOMAS.

PI Canaan E, Croce CM;

DR WPI; 93-214090/26.

P-PSDB; R38470.

PT Detection and treatment of acute leukaemia(s) - using prods.

PT derived from oligo:nucleotide sequences within the ALL-1 gene of

PT chromosome 11

PS Claim 5; Page 29-50; 90pp; English.

CC This sequence represents the cDNA clone of the acute lymphoblastic

CC leukemia (ALL-1) gene of chromosome 11. This gene was isolated by

CC translocation breakpoint mapping. Fragments of the ALL-1 cDNA may

CC be used to identify chromosomal abnormalities within the ALL-1 gene.

CC These fragments may be used in the treatment and diagnosis of human

CC leukemias such as acute lymphocytic, myelomonocytic, monocytic and

CC myelogenous leukemia.

SQ Sequence 1425 BP; 4067 A; 3542 C; 3215 G; 3431 T;

Query Match 77.1%; Score 16.2; DB 1; Length 1425;

Best Local Similarity 85.7%; Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atgaaaaggaaagctaaagt 21

DB 2217 ATGAGAGCAAGAAAGTAAAGT 2197

RESULT 8

ID Q53475/c

AC Q53475;

DT 28-JUN-1994 (first entry)

DE 11q23 chromosome translocation probe MLL 1.5EB.

KW Leukemia; leukemia; MLL gene rearrangement; detection; ss.

OS Homo sapiens.

PN WO9325713-A.

PD 23-DEC-1993.

PF 17-JUN-1992; US-900689.

PR 16-DEC-1992; US-991244.

PA (ARCH-) ARCH DEV CORP.

PI Diaz MO, Rowley JD;

DR WPI; 94-007568/01.

PT Detecting MLL gene rearrangements and translocation - by using

PT nucleic acid probes, for diagnosing leukemia

PS Claim 10; Page 86-87; 136pp; English.

CC The sequence is that of MLL 1.5EB, a probe used in a method for

CC detecting leukemic cells containing 11q23 chromosome translocations.

CC The method is useful in the diagnosis and monitoring of certain

CC types of leukemia using northern and Southern blot analysis and

CC fluorescence in situ hybridisation (FISH). The probe detects

CC rearrangements in dividing leukemic cells and interphase nuclei.

SQ Sequence 1420 BP; 451 A; 323 C; 348 G; 298 T;

Query Match 77.1%; Score 16.2; DB 1; Length 1420;

Best Local Similarity 85.7%; Pred. No. 1.2e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atgaaaaggaaagctaaagt 21

DB 70 ATGAGAGCAAGAAAGTAAAGT 50

RESULT 9

ID Q53477/c

AC Q53477;

DT 28-JUN-1994 (first entry)

DE 11q23 chromosome translocation probe cDNA clone 14P-18B.

KW Leukemia; leukemia; MLL gene rearrangement; detection; ss.

OS Homo sapiens.

PN WO9325713-A.

PD 23-DEC-1993.

PF 17-JUN-1993; US-905857.

PR 17-JUN-1992; US-900689.

PR 16-DEC-1992; US-991244.

PA (ARCH-) ARCH DEV CORP.

PI Diaz MO, Rowley JD;

DR WPI; 94-007568/01.

PT Detecting MLL gene rearrangements and translocation - by using

PT nucleic acid probes, for diagnosing leukemia

PS Claim 13: Page 88-92; 136pp; English.  
 CC The sequence is that of cDNA clone 14P-185, derived from the MLL  
 CC gene and used as a probe in a method for detecting leukaemic cells  
 CC containing 11q23 chromosome translocations. The method is useful  
 CC in the diagnosis and monitoring of certain types of leukaemia  
 CC using northern and Southern blot analysis and fluorescence in  
 CC situ hybridisation (FISH). The probe detects rearrangements in  
 CC dividing leukaemic cells and interphase nuclei.  
 SQ Sequence 4201 BP; 1273 A; 996 C; 989 G; 943 T;

Query Match 77.1%; Score 16.2; DB 1; Length 4201;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atgaagaagaaagctaaagt 21  
 ||| ||||| |||||  
 Db 70 ATGAGAGAGAAAGCTAAAGT 50

RESULT 10  
 Q45339/C  
 ID Q45339 standard; cDNA; 11907 BP.  
 AC Q45339;  
 DT 27-SEP-1994 (first entry)  
 DE cDNA encoding htrx.  
 KW Human; trithorax gene; L01986; diagnosis; treatment;  
 KW immunodeficiency; developmental abnormalities; inherited diseases;  
 KW cancer; acute lymphocytic leukaemia; myelomonocytic leukaemia; ss.  
 OS Homo sapiens.  
 PN W09407302-A.  
 PD 14-APR-1994.  
 PF 24-SEP-1993; U09087.  
 PR 30-SEP-1992; US-954112.  
 PR 13-MAY-1993; US-061376.  
 PA (SALK ) SALK INST BIOLOGICAL STUDIES.  
 PI Djabaki M, Evans GA, Parry P, Selleri L;  
 DR WPI: 94-135206/16.  
 PT Nucleic acid encoding a human trithorax protein - used to  
 PT develop agents for diagnosis and treatment of diseases associated  
 PT with disruption of chromosome II at q23  
 PS Claim 4; page 36-42; 68pp; English.  
 CC In the course of the construction of a physical map of human  
 CC chromosome region 11q23, a region contg. the t(4;11) translocation  
 CC breakpoint was cloned. The cloned DNA encoded a protein homologous  
 CC to the trithorax gene prod. of Drosophila. The gene may be used for  
 CC the diagnosis and treatment of immunodeficiency states,  
 CC developmental abnormalities, inherited diseases or cancers, e.g.  
 CC acute lymphocytic leukaemia or acute myelomonocytic leukaemia.  
 CC Sequence 11907 BP; 3469 A; 2990 C; 2735 G; 2709 T;

Query Match 77.1%; Score 16.2; DB 1; Length 11907;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atgaagaagaaagctaaagt 21  
 ||| ||||| |||||  
 Db 2400 ATGAGAGAGAAAGCTAAAGT 2380

RESULT 11  
 Q75181/C  
 ID Q75181 standard; cDNA; 14255 BP.  
 AC Q75181;  
 DT 18-AUG-1995 (first entry)  
 DE ALL-1 (acute lymphocytic leukaemia-1) cDNA.  
 KW Acute lymphoblastic leukaemia; acute nonlymphoblastic leukaemia;  
 KW chromosomal translocation; abnormality; detection; t(4;11);  
 KW t(9;11); t(11;17); ALL-1; AF-4; AF-9; AF-17; chimeric gene; probe; ds.  
 OS Homo sapiens.  
 PI Key  
 FT exon Location/Qualifiers  
 FT 3901..4032

FT FT /tag- a  
 FT FT /number- 7  
 FT FT /note- "primers for amplifying chimeric sequences  
 FT FT are pref. derived from this region"  
 FT FT 4147..4293  
 FT FT /tag- b  
 FT FT /number- 9  
 FT FT /note- "primers for amplifying chimeric sequences  
 FT FT are pref. derived from this region"  
 FT FT 1..1173  
 FT FT /tag- c  
 FT FT /note- "nucleotides 9353-9696 were found to be  
 FT FT nearly identical to an anonymous sequence  
 FT FT (EST00626) cloned from human foetal brain  
 FT FT cDNA library"

PN W09426930-A.  
 PD 24-NOV-1994.  
 PF 22-APR-1994; U04496.  
 PR 14-MAY-1993; US-062443.  
 PA (UYJE-) UNIV JEFFERSON THOMAS.  
 PI Canaan E, Croce C;  
 DR WPI: 95-006818/01.  
 DR P-PSDB; R66462.  
 PT New acute lymphocytic leukaemia gene prods. - used for the  
 PT diagnosis and treatment of leukaemias, partic. acute  
 PT lymphoblastic or nonlymphoblastic leukaemia  
 PS Claim 65; Page 62-79; 207pp; English.  
 CC The ALL-1 cDNA was obtained from a series of overlapping clones  
 CC spanning 14.7 kb, isolated by screening a human fibroblast library  
 CC and a K562 library. The sequence revealed a single, long ORF  
 CC predicting a protein of approx. 4000 amino acids. The predicted  
 CC amino acid sequence includes regions of homology with the  
 CC cDNA of the trithorax gene which contain zinc-finger domains.  
 CC The ALL-1 gene on chromosome 11 is involved in a series of  
 CC chromosomal translocations which are associated with certain  
 CC leukaemias.  
 SQ Sequence 14255 BP; 4067 A; 3542 C; 3215 G; 3431 T;

Query Match 77.1%; Score 16.2; DB 1; Length 14255;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atgaagaagaaagctaaagt 21  
 ||| ||||| |||||  
 Db 2217 ATGAGAGAGAAAGCTAAAGT 2197

RESULT 12  
 T16330/C  
 ID T16330 standard; DNA; 1420 BP.  
 AC T16330;  
 DT 28-JUL-1997 (revised)  
 DT 06-JUN-1996 (first entry)  
 DE MLL 1.585 fragment for detecting 11q23 chromosome translocations.  
 KW MLL; myeloid; lymphoid; leukaemia; probe; chromosome; translocation;  
 KW mutation; 11q23; lymphoma; diagnosis; monitoring; ss.  
 OS Homo sapiens.  
 PN US5467970-A.  
 PD 30-JAN-1996.  
 PF 17-JUN-1992; 900689.  
 PR 17-JUN-1992; US-900689.  
 PR 16-DEC-1992; US-991244.  
 PR 17-JUN-1993; US-080255.  
 PA (ARCH-) ARCH DEV CORP.  
 PI Diaz MO, Rowley JD;  
 DR WPI: 96-105221/11.  
 PT Detection of 11q23 chromosome translocation(s) - using myeloid/lymphoid  
 PT leukaemia nucleic acid probes, for diagnosis and monitoring of  
 PT leukaemia(s) and lymphoma(s)  
 PS Claim 9; Column 43-44; 47pp; English.  
 CC T16329 and T16330 are human genomic DNA sequences derived from the MLL  
 CC (myeloid/lymphoid leukaemia) gene. The sequences are used as probes

CC for 11q23 chromosome translocations associated with various human  
CC leukaemias. Hybridisation of the probes to aberrant sized DNA  
CC segments is indicative of an 11q23 chromosome translocation. The  
CC probes are useful for the diagnosis and continued monitoring of various  
CC types of leukaemia, partic. myeloid and lymphoid leukaemias and  
CC lymphomas in humans. The probes may also be used for protein prodn.  
CC and hence antibody prodn.  
CC (Revised entry submitted to correct sequence analysis breakdown.)  
SQ Sequence 1420 BP; 451 A; 323 C; 348 G; 298 T;

Query Match 77.1%; Score 16.2; DB 1; Length 1420;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgaagaaggaagctaaagt 21  
||||| ||||| ||||| |||||  
DB 70 ATGAGAGAGAAAGTAAAGT 50

RESULT 13  
T16332/c  
ID T16332 standard; DNA: 4201 BP.

AC T16332;  
DT 29-JUN-1997 (revised)  
DT 06-JUN-1996 (first entry)  
DE MLT. cDNA clone 14-18B for detecting 11q23 chromosome translocations.  
KW MLT. myeloid; lymphoid; leukaemia; probe; chromosome; translocation;  
KW mutation; 11q23; lymphoma; diagnosis; monitoring; ss.  
OS Homo sapiens.  
FT key Location/Qualifiers  
FT cds 2..4201  
FT /tag= a  
FT /note= "product may be used for antibody prodn."

PN US5487970-A.  
PD 30-JAN-1996.  
PF 17-JUN-1992; 900689.  
PR 17-JUN-1992; US-900689.  
PR 16-DEC-1992; US-991244.  
PR 17-JUN-1993; US-080255.  
PA (ARCH-) ARCH DEV CORP.  
PI Diaz MO, Rowley JD;  
DR WPI: 96-105221/11.  
P-PSDB: R92705.  
PT Detection of 11q23 chromosome translocation(s) - using myeloid/lymphoid  
PT leukaemia nucleic acid probes, for diagnosis and monitoring of  
PT leukaemia(s) and lymphoma(s).  
PS Claim 12; Column 43-48; 47pp; English.  
CC T16331 and T16332 are human MLT (myeloid/lymphoid leukaemia) gene  
CC clones 14-7 and 14-18B. The clones are used as probes for 11q23  
CC chromosome translocations associated with various human leukaemias.  
CC Hybridisation of the probes to aberrant sized DNA segments is  
CC indicative of an 11q23 chromosome translocation. The probes are  
CC useful for the diagnosis and continued monitoring of various types  
CC of leukaemia, partic. myeloid and lymphoid leukaemias and lymphomas  
CC in humans. The clones may also be used for protein prodn. (see  
CC R92705 and R92706) and hence antibody prodn.  
CC (Revised entry submitted to correct sequence analysis breakdown.)  
SQ Sequence 4201 BP; 1273 A; 996 C; 989 G; 943 T;

Query Match 77.1%; Score 16.2; DB 1; Length 4201;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 atgaagaaggaagctaaagt 21  
||||| ||||| ||||| |||||  
DB 70 ATGAGAGAGAAAGTAAAGT 50

RESULT 14  
T58840\_0  
WP Sequence split into 6 fragments LOCUS T58840 Accession T58840

WP	Fragment Name	Begin	End
WP	T58840_0	1	110000
WP	T58840_1	100001	210000
WP	T58840_2	200001	310000
WP	T58840_3	300001	410000
WP	T58840_4	400001	510000
WP	T58840_5	500001	580073
ID	T58840 standard; DNA: 580073 BP.		
AC	T58840;		
DT	27-MAR-1997 (first entry)		
DE	Mycoplasma genitalium genome.		
KW	M. genitalium; DNA; DNA gyrase; origin of replication;		
KW	megabase shotgun sequencing method; open reading frame; ORF; ss.		
OS	Mycoplasma genitalium.		
FT	Key Location/Qualifiers		
FT	cds 8552..9184		
FT	/tag= a		
FT	/label= MG006		
FT	/note= "Previously identified as MORF-20076, the		
FT	encoded protein shows 27.59 percentage		
FT	identity to thymidylate kinase (CDC8)		
FT	from Saccharomyces cerevisiae"		
FT	11252..12040		
FT	cds		
FT	/tag= b		
FT	/label= MG009		
FT	/note= "Previously identified as MORF-20078, the		
FT	encoded protein shows 35.43 percentage		
FT	identity to the Bacillus subtilis hypothetical		
FT	protein covered in accession number		
FT	GB:D26185.102"		
FT	12069..12725		
FT	cds		
FT	/tag= c		
FT	/label= MG010		
FT	/note= "Previously identified as MORF-20079, the		
FT	encoded protein shows 25.73 percentage		
FT	identity to DNA primase (dnae) from		
FT	Clostridium acetobutylicum"		
FT	complement (13570..14247)		
FT	cds		
FT	/tag= d		
FT	/label= MG012		
FT	/note= "Previously identified as MORF-20080, the		
FT	encoded protein shows 31.50 percentage		
FT	identity to the ribosomal protein S6		
FT	modification protein (rimk) from Escherichia		
FT	coli"		
FT	complement (14396..15217)		
FT	cds		
FT	/tag= e		
FT	/label= MG013		
FT	/note= "Previously identified as MORF-19823, MORF-20080		
FT	and MORF-20081, the encoded protein shows 33.04		
FT	percentage identity to 5,10-methylene-tetra-		
FT	hydrofolate dehydrogenase (fold) from E. coli"		
FT	17474..19243		
FT	cds		
FT	/tag= f		
FT	/label= MG015		
FT	/note= "Previously identified as MORF-20084, the		
FT	encoded protein shows 32.23 percentage		
FT	identity to transport ATP-binding protein		
FT	(msbA) from E. coli"		
FT	26478..27344		
FT	cds		
FT	/tag= g		
FT	/label= MG023		
FT	/note= "Previously identified as MORF-20092, the		
FT	encoded protein shows 45.96 percentage		
FT	identity to fructose-bisphosphate aldolase		
FT	(tst) from B. subtilis"		
FT	27345..28448		
FT	cds		
FT	/tag= h		
FT	/label= MG024		
FT	/note= "Previously identified as MORF-19826 and		
FT	MORF-20093, the encoded protein shows 46.84		
FT	percentage identity to GTP-binding protein		
FT	from E. coli"		

FT	cds	36987. .38978
FT		/*tag- l
FT		/label= MG032
FT		/note= "Previously identified as MORF-20099, the encoded protein shows 26.82 percentage identity to ATP-dependent nuclease (adda) from B. subtilis"
FT	cds	39242. .39904
FT		/*tag- j
FT		/label= MG033
FT		/note= "Previously identified as MORF-20100, the encoded protein shows 35.90 percentage identity to glycerol uptake facilitator (glpf) from B. subtilis"
FT	cds	complement (39873. .40514)
FT		/*tag- k
FT		/label= MG034
FT		/note= "Previously identified as MORF-20101, the encoded protein shows 48.13 percentage identity to thymidylate kinase (tdk) from B. subtilis"
FT	cds	40543. .41787
FT		/*tag- l
FT		/label= MG035
FT		/note= "Previously identified as MORF-20102, the encoded protein shows 30.71 percentage identity to histidyl-tRNA synthetase (hisS) from Mycobacterium leprae"
FT	cds	complement (44751. .46277)
FT		/*tag- m
FT		/label= MG038
FT		/note= "Previously identified as MORF-20105, the encoded protein shows 46.83 percentage identity to glycerol kinase (gpk) from E. coli"
FT	cds	complement (46268. .47422)
FT		/*tag- n
FT		/label= MG039
FT		/note= "Previously identified as MORF-19831 and MORF-20106, the encoded protein shows 43.20 percentage identity to glycerol-3-phosphate dehydrogenase (GUT2) from S. cerevisiae"
FT	cds	49377. .49643
FT		/*tag- o
FT		/label= MG041
FT		/note= "The encoded protein shows 48.86 percentage identity to phosphohistidinoprotein-hexose phosphotransferase (ptsh) from Mycoplasma capricolum"
FT	cds	50060. .51520
FT		/*tag- p
FT		/label= MG042
FT		/note= "Previously identified as MORF-19832 and MORF-20108, the encoded protein shows 41.92 percentage identity to spermidine/ putrescine transport ATP-binding protein (potA) from E. coli"
FT	cds	51925. .52382
FT		/*tag- q
FT		/label= MG043
FT		/note= "Previously identified as MORF-20110, the encoded protein shows 26.51 percentage identity to spermidine/putrescine transport system permease protein (potB) from E. coli"
FT	cds	52366. .53220
FT		/*tag- r
FT		/label= MG044
FT		/note= "Previously identified as MORF-20111, the encoded protein shows 29.45 percentage identity to spermidine/putrescine transport system permease protein C (potC) from E. coli"
FT	cds	54658. .55605
FT		/*tag- s

[illegible]

```

FH Key Location/Qualifiers
FT exon 157..300
FT /tag= a
FT /number= 5
FT /note= "Approximate position, determined from data
FT listed in Table 4 of the specification and
FT comparison with cDNA sequence in T51254"
PN WO9703192-A2.
PD 30-JAN-1997.
PE 05-JUL-1996: U11386.
PR 14-AUG-1995: US-002328.
PR 07-JUL-1995: US-000956.
PR 28-JUL-1995: US-001675.
PR 11-AUG-1995: US-002174.
PA (DARW-) DARWIN MOLECULAR CORP.
PA (GEHO-) GEN HOSPITAL CORP.
PA (VAME-) VA MEDICAL CENT.
PI Bird TD, Galas DJ, Levy-Lahad E, Mulligan J, Schellenberg GD;
PI Tanzi RE, Wasco W;
PI WPI; 97-119048/11.
PT New Alzheimer's disease related gene, AD4 - used to develop prods.
PT for detecting pre-disposition to or for diagnosis, prevention or
PT treatment of Alzheimer's disease
PS Claim 46; Fig 16; 83pp; English.
CC A genetically isolated group of families with autosomal dominant
CC early-onset Alzheimer's Disease (AD) (the Voilga German kindreds) has
CC been studied and initial mapping analyses have predicted the AD4
CC locus resides on chromosome 1. A 2.5 kb fragment of DNA from the
CC AD4 gene was amplified from YAC 921d12 DNA and was radiolabelled by
CC nick translation. A human genomic clone hybridising to this fragment
CC was identified in a commercially available library. The clone carried
CC the entire AD4 gene and was sheared and subcloned into M13mp18 vector
CC for sequence analysis. The sequences of the 12 exons and the flanking
CC intron sequences of the AD4 gene were determined. The present sequence
CC includes exon 5. Mutations in the AD4 gene, identified in Voilga German
CC kindreds, are associated with Alzheimer's Disease, especially a
CC mutation at codon 141 which replaces Asn by Ile. Detection of mutant
CC AD4, for example using antibodies specific for the protein or using
CC nucleic acid probes specific for the mutant gene, provides a means of
CC diagnosing Alzheimer's disease.
SQ Sequence 1438 BP; 319 A; 398 C; 346 G; 375 T;

```

```

Query Match 77.1%; Score 16.2; DB 1; Length 1438;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 atgaagaagaagctaaagt 21
   1 ||||||| |||||
DB 964 AGGAAAAGCAAACTAAAAAT 944

```

Search completed: August 18, 1999, 17:18:31  
Job time: 3263 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 17:04:38 ; Search time 767.16 Seconds  
(without alignments)  
53.996 Million cell updates/sec

Title: US-09-004-395-3

Perfect score: 21

Sequence: 1 atgaaaagaaagctaaagt 21

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database: EST\*

1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18.4	87.6	211	34	AA514077	AA514077 MC2AS032.
2	17.8	84.8	428	21	R19941	R19941 Y930F02.r1
3	17.8	84.8	397	21	T87926	T87926 Y608C06.r1
4	17.8	84.8	401	23	D62554	D62554 HUM30B06B
5	17.8	84.8	355	23	D63016	D63016 HUM352D07B
6	17.8	84.8	296	25	N73600	N73600 za51d06.s1
7	17.8	84.8	217	37	AA700025	AA700025 z169d12.s
8	17.8	84.8	414	40	AA987196	AA987196 G192d11.s
9	17.8	84.8	592	43	AA179333	AA179333 EST223029
10	17.8	84.8	491	45	AA171083	AA171083 ta07f09.x
11	17.8	84.8	511	47	AA147812	AA147812 tm52f03.x
12	17.8	84.8	506	48	AA1589168	AA1589168 tf81b04.x
13	17.8	84.8	442	50	AA1694973	AA1694973 we44b08.x
14	17.4	82.9	395	21	T95759	T95759 ye40d12.r1
15	16.8	80.0	467	24	N29345	N29345 yw85c10.s1
16	16.8	80.0	433	27	AA009986	AA009986 z16g10.r
17	16.8	80.0	581	34	AA527160	AA527160 n107c11.s
18	16.8	80.0	126	41	AU013107	AU013107 AU013107
19	16.8	80.0	370	50	AA025762	AA025762 AV025762
20	16.4	78.1	463	21	R18813	R18813 Y922C02.r1
21	16.4	78.1	466	25	N72384	N72384 yv38g12.r1
22	16.4	78.1	433	26	W68463	W68463 zd36e07.r1
23	16.4	78.1	389	26	W92378	W92378 zd99g12.s1
24	16.4	78.1	400	28	AA058842	AA058842 z164b04.s
25	16.4	78.1	527	29	AA164406	AA164406 z097h11.r
26	16.4	78.1	428	29	AA164407	AA164407 z097h11.r
27	16.4	78.1	463	33	AA406425	AA406425 zv12c11.r
28	16.4	78.1	436	33	AA410434	AA410434 zv12c11.s
29	16.4	78.1	435	33	AA451689	AA451689 zx44c07.s
30	16.4	78.1	485	34	AA482398	AA482398 z134b06.r
31	16.4	78.1	464	34	AA482544	AA482544 z134b06.r
32	16.4	78.1	477	34	AA488811	AA488811 aa54f05.r
33	16.4	78.1	582	35	AA584310	AA584310 nm79g01.s
34	16.4	78.1	544	37	AA678638	AA678638 ah07g11.s
35	16.4	78.1	403	37	AA723450	AA723450 z969c04.s
36	16.4	78.1	352	38	AA738416	AA738416 nx19f10.s
37	16.4	78.1	332	39	AA843216	AA843216 ak06f05.s
38	16.4	78.1	425	39	AA857920	AA857920 o169c01.s
39	16.4	78.1	520	40	AA917064	AA917064 o146g10.s
40	16.4	78.1	335	40	AA969841	AA969841 op15f07.s
41	16.4	78.1	340	40	AA974239	AA974239 cq10c09.s
42	16.4	78.1	483	40	AA983295	AA983295 cq56e07.s
43	16.4	78.1	502	40	AA992914	AA992914 oc92c05.s
44	16.4	78.1	433	41	AI039948	AI039948 ox49f08.x
45	16.4	78.1	444	49	AI653051	AI653051 wb42h03.x

## ALIGNMENTS

RESULT 1  
AA514077  
LOCUS AA514077 211 bp mRNA  
DEFINITION MC2AS032.ACS S.mansoni cercarial Lambda zap Schistosoma mansoni  
CDNA clone AS-32 5', mRNA sequence.  
ACCESSION AA514077  
NID 92252933  
VERSION AA514077.1 GI:2252933

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1657  
High quality sequence strops: 235 Source: IMAGE Consortium, LLNL

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [estewatson.wustl.edu](mailto:estewatson.wustl.edu)  
Insert Size: 641

High quality sequence stops: 179. Source: IMAGE Consortium, LBNL  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.



Insert Length: 641 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 179.  
Location/Qualifiers

## FEATURES

source

1..397  
/organism="Homo sapiens"  
/db\_xref="GDB:485419"  
/db\_xref="taxon:9606"  
/clone="IMAGE:117130"  
/clone\_1lb="Stratagene Lung (#937210)"  
/sex="male"  
/dev\_stage="72 years"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: Lung; Vector: pBluescript SK-; site\_1:  
ECORI; site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. normal lung. Average insert size: 1.0 kb;  
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAAATCGGCGACGAG  
3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

## BASE COUNT

107 a 75 c 70 g 138 t 7 others

ORIGIN

Query Match 84.8% Score 17.8: DB 21: Length 397;  
Best Local Similarity 90.5% Pred. No. 5.7e+02;  
Matches 19: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

QY 1 atgaaaggaaagctaaagt 21  
|||||

Db 220 ATGAAAGGAATGCTAGAGT 200

RESULT 4  
D62554/c 401 bp mRNA EST 29-AUG-1995  
LOCUS HUM300B06b clonetech human aorta polyA+ mRNA (#6572) Homo sapiens  
DEFINITION cDNA clone GEN-300B06 5', mRNA sequence.  
ACCESSION D62554  
NID 9966328  
VERSION D62554.1 GI:966328  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

AUTHORS

1 (bases 1 to 401)  
Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,  
Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H.,  
Takahashi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y.,  
Maekawa, H., Shin, S. and Nakamura, Y.  
Fujiwara et al. (1995)  
Unpublished (1995)  
On Apr 14, 1993 this sequence version replaced gi:693411.

## TITLE

JOURNAL  
Unpublished (1995)  
On Apr 14, 1993 this sequence version replaced gi:693411.

## COMMENT

Contact: Tsutomu Fujiwara  
Otsuka GEN Research Institute  
Otsuka Pharmaceutical Co., Ltd  
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan  
Tel: 0886-65-2888  
Fax: 0886-37-1035.

## FEATURES

source

Location/Qualifiers  
1..401  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="1p"  
/clone="GEN-300B06"  
/clone\_1lb="Clontech human aorta polyA+ mRNA (#6572)"

## BASE COUNT

115 a 64 c 61 g 149 t 12 others

ORIGIN

Query Match 84.8% Score 17.8: DB 21: Length 401;  
Best Local Similarity 90.5% Pred. No. 5.7e+02;  
Matches 19: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

QY 1 atgaaaggaaagctaaagt 21  
|||||

Db 137 ATGAAAGGAATGCTAGAGT 117

RESULT 5  
D63016/c 355 bp mRNA EST 29-AUG-1995  
LOCUS HUM352D07b clonetech human aorta polyA+ mRNA (#6572) Homo sapiens  
DEFINITION cDNA clone GEN-352D07 5', mRNA sequence.  
ACCESSION D63016  
NID 9966790  
VERSION D63016.1 GI:966790  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

AUTHORS

1 (bases 1 to 355)  
Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,  
Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H.,  
Takahashi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y.,  
Maekawa, H., Shin, S. and Nakamura, Y.  
Fujiwara et al. (1995)  
Unpublished (1995)  
On Apr 14, 1993 this sequence version replaced gi:716676.

## TITLE

JOURNAL  
Unpublished (1995)  
On Apr 14, 1993 this sequence version replaced gi:716676.

## COMMENT

Contact: Tsutomu Fujiwara  
Otsuka GEN Research Institute  
Otsuka Pharmaceutical Co., Ltd  
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan  
Tel: 0886-65-2888  
Fax: 0886-37-1035.

## FEATURES

source

Location/Qualifiers  
1..355  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="932A02: 2: 2g24.3-2g31.3 (2g31)"  
/clone="GEN-352D07"  
/clone\_1lb="Clontech human aorta polyA+ mRNA (#6572)"

## BASE COUNT

105 a 62 c 57 g 131 t

ORIGIN

Query Match 84.8% Score 17.8: DB 23: Length 355;  
Best Local Similarity 90.5% Pred. No. 5.8e+02;  
Matches 19: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

QY 1 atgaaaggaaagctaaagt 21  
|||||

Db 84 ATGAAAGGAATGCTAGAGT 64

RESULT 6  
N73600 296 bp mRNA EST 19-MAR-1996  
LOCUS za51d06.s1 Soares fetal liver spleen JNFLS Homo sapiens cDNA clone  
DEFINITION IMAGE:296075 3' similar to contains Alu repetitive element; mRNA  
sequence.  
ACCESSION N73600  
NID 91230885  
VERSION N73600.1 GI:1230885  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

AUTHORS

1 (bases 1 to 296)  
Haller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, D., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and



Seq primer: -40m13 fwd. ET from Amerisham  
High quality sequence stop: 407.

FEATURES  
source  
1. .414  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="9"  
/clone="IMAGE:1603292"  
/clone\_1ib="NCI\_CGAP\_Lu5"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"

BASE COUNT 150 a 63 c 75 g 126 t  
ORIGIN  
Query Match 84.8%; Score 17.8; DB 40; Length 414;  
Best Local Similarity 90.5%; Pred. No. 5.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaaagt 21  
|||||  
Db 75 ATGAAAGGAACTGCTGAGAGT 95

RESULT 9  
A1179333 592 bp mRNA EST 20-JAN-1999  
LOCUS EST223029 Normalized rat spleen, Bento Soares Rattus sp. cDNA clone  
DEFINITION RSPCF39.3' end, mRNA sequence.  
ACCESSION A1179333  
NID 93729971  
VERSION A1179333.1 GI:3729971  
KEYWORDS EST.  
SOURCE Rattus sp.  
ORGANISM Rattus sp.  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE Lee, N.H., Glodet, A., Chandra, I., Mason, T.M., Quackenbush, J.,  
Kerlavage, A.R. and Adams, M.D.  
Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat  
Gene Index  
TITLE Unpublished (1998)  
JOURNAL On Aug 21, 1998 this sequence version replaced.  
COMMENT

Contact: Lee, NH  
ATCC  
The Institute for Genomic Research  
9712, Medical Center Drive, Rockville, MD 20850, USA  
Tel: (301)-838-3529  
Fax: (301)-838-0208  
Email: nhlee@igrr.org  
Seq primer: M13-21.  
Location/Qualifiers

FEATURES  
source  
1. .592  
Location/Qualifiers  
/organism="Rattus sp."  
/db\_xref="ATCC (lnhost):2035046"  
/db\_xref="taxon:10118"  
/clone="RSPCF39"  
/clone\_1ib="Normalized rat spleen, Bento Soares"  
/note="Organ: spleen; Vector: pRT73Pac; Site\_1: EcoRI;  
Site\_2: NotI"

BASE COUNT 145 a 152 c 171 g 124 t  
ORIGIN

Query Match 84.8%; Score 17.8; DB 43; Length 592;  
Best Local Similarity 90.5%; Pred. No. 5.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaaagt 21  
|||||  
Db 102 ATGAAAGGAAAGCTGAAAGT 122

RESULT 10  
A1371083 491 bp mRNA EST 13-APR-1999  
LOCUS ta07109.x1 NCI\_CGAP\_Brn23 Homo sapiens cDNA clone IMAGE:2043401 3',  
DEFINITION mRNA sequence.  
ACCESSION A1371083  
NID 94149836  
VERSION A1371083.1 GI:4149836  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 491)  
NCBI/INSD-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute / National Institute of Neurological  
TITLE Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BNGAP), Tumor Gene Index  
Unpublished (1998)  
JOURNAL On Jan 17, 1998 this sequence version replaced gi:2043654.  
COMMENT

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfield M.D.,  
Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbp/IMAGE/IMAGE.html

Insert length: 786 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 417.  
Location/Qualifiers

FEATURES  
source  
1. .491  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="7q32; D7S686-D7S640"  
/clone="IMAGE:2043401"  
/clone\_1ib="NCI\_CGAP\_Brn23"  
/tissue\_type="glioblastoma (pooled)"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pRT73D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer (5'  
TGTACCAATCTGAAAGTGGAGCGCCGCAATCTTTTCTTTTCTTTT  
T 3'); double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pRT73 vector.  
Library is normalized, and was constructed by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 183 a 83 c 73 g 152 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 45; Length 491;  
Best Local Similarity 90.5%; Pred. No. 5.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaaagt 21  
|||||  
DB 450 ATGAAAGGAGATGCTAGAAGT 470

RESULT 11  
A1478812 511 bp mRNA EST 14-APR-1999  
LOCUS tlm2f03.x1 NCI\_CGAP\_K1d11 Homo sapiens cDNA clone IMAGE:2161757 3',  
DEFINITION mRNA sequence.  
ACCESSION A1478812  
NID 94373625  
VERSION A1478812.1 GI:4373625  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 511)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189409.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskeluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bdrrp/image/image.html

Insert Length: 898 Std Error: 0.00  
Seq primer: -40UP from G1bco  
High quality sequence stop: 468.

FEATURES  
Location/Qualifiers  
1..511  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="890D10; 5: 5q23.2-5q31.1"  
/clone="IMAGE:2161757"  
/clone\_lib="NCI\_CGAP\_K1d11"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_K1d3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(clones 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 195 a 84 c 73 g 159 t  
ORIGIN

Query Match 84.8%; Score 17.8; DB 47; Length 511;  
Best Local Similarity 90.5%; Pred. No. 5.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaaagt 21  
|||||  
DB 442 ATGAAAGGAGATGCTAGAAGT 462

RESULT 12  
A1589168  
NID

LOCUS A1589168 506 bp mRNA EST 12-MAY-1999  
DEFINITION t81h04.x1 NCI\_CGAP\_Brn23 Homo sapiens cDNA clone IMAGE:2105719 3',  
ACCESSION A1589168  
NID 94598216  
VERSION A1589168.1 GI:4598216  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 506)  
AUTHORS NCI/NIHDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BTCAP), Tumor Gene Index  
JOURNAL Unpublished (1998)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948176.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfield M.D.,  
Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bdrrp/image/image.html

Insert Length: 819 Std Error: 0.00  
Seq primer: -40UP from G1bco  
High quality sequence stop: 412  
POLYA-No.

FEATURES  
Location/Qualifiers  
1..506  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2105719"  
/clone\_lib="NCI\_CGAP\_Brn23"  
/lab\_host="DH10B"  
/tissue\_type="glioblastoma (pooled)"  
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer (5'  
TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTTCTTTT  
T 3'); double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
Library is normalized, and was constructed by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 194 a 83 c 73 g 156 t  
ORIGIN

Query Match 84.8%; Score 17.8; DB 48; Length 506;  
Best Local Similarity 90.5%; Pred. No. 5.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 atgaaaggaagctaaagt 21  
|||||  
DB 438 ATGAAAGGAGATGCTAGAAGT 458

RESULT 13  
A1694973 442 bp mRNA EST 03-JUN-1999  
LOCUS w644h08.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:2343999 3',  
DEFINITION mRNA sequence.  
ACCESSION A1694973  
NID 94982873

```

VERSION      AF1694973.1  GI:4982873
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE    1 (bases 1 to 442)
AUTHORS      NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      On Jun 22, 1998 this sequence version replaced gi:3246995.

Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNI at:
www-bio.llnl.gov/dbp/IMAGE/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 367.
Location/Qualifiers
1..442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2343999"
/clone_lib="NCI-CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: Lung; Vector: pRTT3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI-CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneids
141920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT   155 a      70 c      80 g      137 t
ORIGIN
Query Match      84.8%; Score 17.8; DB 50; Length 442;
Best Local Similarity 90.5%; Pred. No. 5.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY      1 atgaaaggaagcctaaagt 21
      |||||||
Db      83 ATGAAAGGAATGCTGAAAGT 103

RESULT      14
LOCUS       T95759      395 bp      mRNA      EST      27-MAR-1995
DEFINITION ye04d12.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:120215 5', mRNA sequence.
ACCESSION   T95759
NID         9734383
VERSION     T95759.1  GI:734383
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE    1 (bases 1 to 395)
AUTHORS      Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maita, M.,

```

```

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 1325
High quality sequence stops: 298 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1325 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 298.
Location/Qualifiers
1. 395
/organism="Homo sapiens"
/db_xref="GDB:472760"
/db_xref="taxon:9606"
/clone="IMAGE:120215"
/clone_1lb="Soares fetal liver spleen JNFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker. Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAGAGATTAAATTAAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonalido."

BASE COUNT 97 a 68 c 72 g 153 t 5 others
ORIGIN

Query Match 82.9% Score 17.4 DB 21 Length 395:
Best Local Similarity 94.7% Pred. No. 8.1e+02:
Matches 18: Conservative 0; Mismatches 1; Indels 0; Gaps 0.

Oy 2 tgaagaaggaagcctaaag 20
| ||||| ||||| |||||
Db 285 TAAAGAAGAAAGCTAAAG 303

RESULT 15
N29345 467 bp mRNA EST 05-JAN-1996
LOCUS N29345
DEFINITION y085c10.s1 Soares-Jlacentis_8c09weeks_2NDHP019W Homo sapiens cDNA
Clone IMAGE:259026 3', mRNA sequence.
ACCESSION N29345
MID g1147865
VERSION N29345.1 GI:1147865
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 467)
Haller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maita, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Woldmann, P. and
Wilson, R.
The WashU-Merck EST project
Unpublished (1995)
On May 18, 1995 this sequence version replaced gi:810992.

```

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [ced@wustl.edu](mailto:ced@wustl.edu)

Email: [estewatson.wustl.edu](mailto:estewatson.wustl.edu)  
High quality sequence stops: 307  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Seq primer: m13 -40 forward  
High quality sequence stop: 307.

## FEATURES

**Source**

```

/organism="Homo sapiens"
/db_xref="GDB:3888732"
/db_xref="taxon:9606"
/clone_1lib="Soares.placenta_8to9weeks_2NBHP8to9w"
/dev_stage="two placentaes: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker: Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(dT) primer [5'
TGTTACCATCTGTAAGGAGGCGCGCGCATATTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Faúma Bonaldo."

```

BASE COUNT	178 a	81 c	69 g	136 t	3 others
ORIGIN					

Query Match	Score 16.84	DB 24	length 467
Best Local Similarity	85.7%	Pred. No. 1,3e+03	
Matches 18; Conservative	0; Mismatches 3;	Indels 0;	Gaps 0;
Qy	1 atgaaaggaagcaagctaaagt	21	
Db	443 atgaaaggaatgcnnagangt	463	

Search completed: August 18, 1999, 17:04:42  
Job time: 2788 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 17:17:01 ; Search time 543.6 Seconds

(without alignments)  
111.159 Million cell updates/sec

Title: US-09-004-395-4

Perfect score: 19  
Sequence: 1 gatgattacagagaggt 19

Scoring table: IDENTITY\_NUC

Searched: 679419 segs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_v1:\*  
18: em\_fun:\*  
19: em\_htg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_v1:\*  
34: gb\_hg1:\*  
35: gb\_hg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pt4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	19	100.0	13271	2	AE001168	Borrelia

2	19	100.0	1655	2	BU62900	U62900 Borrelia bu
3	17	89.5	175001	34	HSB255N20	AL078632 Homo sapi
4	17	89.5	188759	34	HSB255N20	AL050312 Homo sapi
5	16.4	86.3	114600	10	HS477H23	AL033538 Human DNA
6	16.4	86.3	44665	10	HSB107C2	AL008724 Human gen
7	16.4	86.3	226295	34	HS353E16	AL031591 Homo sapi
8	16.4	84.2	125698	10	HS692C8	AL034561 Human DNA
9	15.8	83.2	3105	1	PIP404BCN	M14481 Plasmid pip
10	15.8	83.2	10206	1	PIP404CG	M32882 Plasmid pip
11	15.8	83.2	10058	2	U67501	U67501 Methanococc
12	15.8	83.2	2418	4	CSU23186	U23186 Crocalus sc
13	15.8	83.2	127583	9	HS941P9	Z95331 Human DNA
14	15.8	83.2	156791	10	HSBU86OP4	AL049594 Human DNA
15	15.8	83.2	87857	11	AC005895	AC005895 Homo sapi
16	15.8	83.2	42948	34	AC006622	AC006622 Caenorhab
17	15.8	83.2	110000	34	CEY116A8.3	Continuation (4 of
18	15.8	83.2	110000	34	CEY116A8.4	Continuation (5 of
19	15.8	83.2	110000	34	CEY43D4.0	AL035066 Caenorhab
20	15.8	83.2	143283	34	HS620E11	AL031667 Homo sapi
21	15.8	83.2	188500	34	HSDJ137K2	AL049820 Homo sapi
22	15.8	83.2	233369	35	AC000353	AC000353 Homo sapi
23	15.8	83.2	843	36	PFALSA1A	LA0908 Plasmodium
24	15.8	83.2	843	36	PFALSA1B	LA0909 Plasmodium
25	15.8	83.2	843	36	PFALSA1C	LA0884 Plasmodium
26	15.8	83.2	843	36	PFALSA1D	LA0886 Plasmodium
27	15.8	83.2	843	36	PFALSA1E	LA0887 Plasmodium
28	15.8	83.2	843	36	PFALSA1F	LA0888 Plasmodium
29	15.8	83.2	5970	36	PFALSA1G	LA0892 Plasmodium
30	15.8	83.2	38749	37	CELRI60	LA0893 Plasmodium
31	15.8	83.2	843	37	PFALSA1H	LA0836 Plasmodium
32	15.8	83.2	843	37	PFALSA1I	LA0837 Plasmodium
33	15.8	83.2	843	37	PFALSA1J	M64227 pig prolyl
34	15.8	83.2	843	37	PFALSA1K	A44406 Sequence 3
35	15.8	83.2	843	37	PFALSA1L	A44407 Sequence 4
36	15.8	83.2	843	37	PFALSA1M	I19547 Sequence 3
37	15.8	83.2	843	37	PFALSA1N	I19548 Sequence 4
38	15.8	83.2	843	37	PFALSA1O	AL035263 S.pombe c
39	15.4	81.1	2495	3	PIGPREP	D50493 Flsiston yea
40	15.4	81.1	4027	5	A44406	
41	15.4	81.1	4027	5	A44407	
42	15.4	81.1	4027	5	I19547	
43	15.4	81.1	4027	5	I19548	
44	15.4	81.1	42391	7	SPBC776	
45	15.4	81.1	3644	7	YSPSK1PK	

#### ALIGNMENTS

RESULT 1  
AE001168  
LOCUS 13271 bp DNA BCT 15-DEC-1997  
DEFINITION Borrelia burgdorferi (section 54 of 70) of the complete genome.  
ACCESSION AE001168 AE000783  
NID 92688598  
VERSION AE001168.1 GI:2688598  
KEYWORDS  
SOURCE  
ORGANISM  
Borrelia burgdorferi  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
burgdorferi group.  
1 (bases 1 to 13271)  
REFERENCE  
AUTHORS Fraser,C.M., Castens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,  
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
Gwin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,  
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,  
Salzberg,S., Hanson,M., van Vugt,R., Palmer,N., Adams,M.D.,  
Gocayne,J.D., Weidman,J., Uterback,T., Matthey,L., McDonald,L.,  
Artlich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,  
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
Genomic sequence of a Lyme disease spirochaete, Borrelia  
burgdorferi  
JOURNAL Nature 390 (6660), 580-586 (1997)

MEDLINE  
98065943  
2 (bases 1 to 13271)  
Fraser,C.M., Castiens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,  
AUTHORS Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
Gwin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,  
Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,  
Salzberg,S., Hanson,M., van Vugt,R., Palmer,N., Adams,M.D.,  
Gocayne,J.D., Weidman,J., Uterback,T., Matthey,L., McDonald,L.,  
Artach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,  
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
TITLE  
Submitted (12-Dec-1997) The Institute for Genomic Research, 9712  
JOURNAL Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
Location/Qualifiers  
1..13271  
/organism="Borrelia burgdorferi"  
/db\_xref="taxon:139"  
66..1100  
/gene="BB0668"  
66..1100  
/note="similar to GP:1575447 percent identity: 98.83;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="flagellar filament outer layer protein (flaA)"  
/protein\_id="AAC67025.1"  
/db\_xref="PID:g268608"  
/db\_xref="GI:268608"  
/translation="MYMKRKAASLFFLLSTVLPAGQTDGLAEGSKRAEGEVLVD  
AELADPSSTRDLNVDVYSGASGIVKPEDVDVGINNWSVLLTPSARLQAYK  
NSVAVAVAKSESKRYAGDTILGVAVLPSPYSOSGAMTPPEKIPYSGESGNFLGK  
GLINDIKTEIKVSVSLGYEIDLEVLFEEDMGEMAYSMGTLLKFKMADLINSNP  
YIPNISRRIKDDVPNPYPLASSKMFKAFFRVSKSHSKSEONFIYVNDLRYLYOKLSY  
SIDSIDSESVFKVYETSGTESLRKLAKHEFKRLKREKISMESGFQNFVKIES  
EKPESSPKN"  
1168..3762  
/gene="BB0669"  
1168..3762  
/note="similar to PID:1113815 SP:Q44737 GB:AE00783  
percent identity: 99.31; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis histidine kinase (cheA-2)"  
/protein\_id="AAC67024.1"  
/db\_xref="PID:g268607"  
/db\_xref="GI:268607"  
/translation="MEILDLENELLGVFEAEQNLVDILEENINSLDEDPNNSDTID  
EIFRAHRLKGSASLIDMMLSDFTFIYEDVDALRDGKVINNDVLLSLSDVIR  
EMALRIDGKYLNDISLKSLSKQFLVIDOTFKRFGNSIKNNPCLSDSEIR  
EGIGIGKQVLRISVENSNSNSVSGSLKILNLLKNGSVLHRTIPKREOILDEKFK  
RVDTYLLISDIEGKKSLSLNLISYIYDENRKEELKLADEIKYVDLSNFVLA  
DNFDEDEISDLLEVQKLFKRVDFKDNPAATISGLQMLSLKIRKSID  
DSSELLAKFEDEVLYLISNTSESIKKNLDPVHSEFIKKNVNLKSLSVKRE  
DEAPFEKKNKIKNSPISVNLRIDSKIDYILNYSVAIVSKSYNOINSEMDIEIR  
NEAFYVDOESFORNFLDLKIVFKDGLTLEDESHINSISKFMKALKADISELR  
NSPFLQNFKMTSGRLSDITDLHESYLKRMPLPSNIFSRPRVYVDLKKIKLTY  
NLKMEGETELDKSYIDLVPPLMHCYNSDGHGLEYEERKRSKAGATILBAK  
EGNVYISIEDDGIIDPKVIRKLIENGITKEDVYSDFELINLIFAPGSTAVQVT  
DLGSGVGLDVYKRSIKELNGTILVESIGITFKIKPLPLVLIQGLKVSSEY  
VILPNNVLEFHRITEHDIKLENYHEVNLDEDEVSVRLDKLFNITRDSLLIEKFLI  
VNTSNMKIATVDSILGEDFVYVPIKDKSSSAGIYGATTLGNGKVVLLIDVFKFL  
DLQDKTKR"

gene  
CDS

3769..5169  
/gene="BB0670"  
3769..5169  
/gene="BB0670"  
/note="similar to GP:1881576 percent identity: 98.93;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

/product="purine-binding chemotaxis protein (cheW-3)"  
/protein\_id="AAC67023.1"  
/db\_xref="PID:g268606"  
/db\_xref="GI:268606"  
/translation="MQIKETYPGPKILDDKNSKLNPNDFKVPSELSGDHLYDIM  
QVEIRKSSFTYPPNAKKYVAGIDNIRGITPIIDIRINFNLEFNKKOJEDIMVLKN  
EDLLIGIVKINNVSIDSSLIDPPHVLSDOSLNYIGVDYNEKLYLILNWKI  
FNGEERKLKPGQNFVEKEDFVSDCDNLDILSKNDSEFNKVSXKNDGNSTLNN  
TAAFNLEIKNNILKYSFNASLVNDVLELVYGFNVDVYLPJCSDFLNEFYKSSG  
NMGADCLEEFKNELVSRLSNMNNLSIFNVLEIGGSGKETMALNALSEYVYKPE  
KTAINDLSKVETSRVSESEIGISFISRSNPSPOSGVYKFKSEITLSNLFVYS  
DALESFDPNLGAWTFKDYVLCFLDSKQIILINILASKAIKGLVLDNDELKNNDY  
FIKESKYFNLTKRV"  
5181..5666  
/gene="BB0671"  
5181..5666  
/gene="BB0671"  
/note="similar to GP:1408273 percent identity: 98.76;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis operon protein (cheX)"  
/protein\_id="AAC67022.1"  
/db\_xref="PID:g268605"  
/db\_xref="GI:268605"  
/translation="MRDYIEPLDAASVLRDMLVENVEMGKPGKLSINOKIGVS  
VIVLAGSVSGSLIIDMDIEFLVASKLMEFEYDDDEDETEKEMVATLVEGNILA  
GNFVTLHAKGFVDIPPAFVIGENKINKSGEALIVFSLPDGKILVNIARER  
V"  
5706..6146  
/gene="BB0672"  
5706..6146  
/gene="BB0672"  
/note="similar to GP:1408274 percent identity: 97.95;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis response regulator (cheY-3)"  
/protein\_id="AAC67021.1"  
/db\_xref="PID:g268604"  
/db\_xref="GI:268604"  
/translation="MTQKTTIADSSSKPQINVDGIPPNVLVDSDVTVQOLTOI  
VITSEGFNIDIPADGEAVIRKKNHPNIDIVLDTMPMDGITCLSNIMEDKNAR  
VIMSALGKBDLVKDCILKGAKTFTVPLRAVLOKMSVFK"  
complement(6187..6702)  
/gene="BB0673"  
complement(6187..6702)  
/gene="BB0673"  
/note="similar to GP:1408275 percent identity: 97.56;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC67020.1"  
/db\_xref="PID:g268603"  
/db\_xref="GI:268603"  
/translation="MKIQIILMLALDPLNARLDLSIEKRADEIRKYSYNLLI  
EKETVNPPESEIKNIYKLEHFKVKSIMLNKNYSLNYSKAEKYLQSELDIKK  
FKRIKIKININGIFRSHSLITKRFYLELLEINNAPELKFNLNITYFLKNDK  
ISNEMIFPRE"  
complement(6699..7745)  
/gene="BB0674"  
complement(6699..7745)  
/gene="BB0674"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0674"  
/protein\_id="AAC67027.1"  
/db\_xref="PID:g268610"  
/db\_xref="GI:268610"  
/translation="MYKKSIFLKAIIISIFELLLELSIIIFLPYKIRFALLIFGP



LEPTIFIFLYKITKAYLSQRLEIYVRNNLFEDIIHCLIPLAFTSYQKNITVAHE  
TINPIMLSLFKRLRLRNDLIIIEIYNSKKNLILIAFARFMSLLIPETFI  
IISSEKIVNISPEKOEENIKINISINERKAYIKERYPILIKEDODIYKSDIEIV  
YSPSEKVIEMEKTEKFIIDYLOKSDSLIGLFTLPASFTIPLNFKPRASFL  
NPILIMTKLIODPLEYKRIQIPTLSEKYTELAKEFNLLKKNLSKRSKRIPLFI  
EKVKIINKMQEIK"  
gene complement(7739. .8602)  
/gene="B80675"  
/complement(7739. .8602)  
/gene="B80675"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region B80675"  
/protein\_id="AAC67026.1"  
/db\_xref="PID:g288609"  
/db\_xref="GI:2688609"  
/translation="MNTKPYFYLHYLFNHSIKSISAEKEIEILNLYKKNRTIA  
TFIKNDESEIKDIQYKDKTDIMTPVLSGEALDPNIVKPLFSEKELKNDLNI  
FNFVKNVSLRKEEFYNTISNGYITFYINKLEBKNSYIYLIOKNALYSDDII  
KNYIKILLKLVIVIKYCFEKGTELTTRKNIESKATSDNDFDEKATLIISEPK  
YETLQTPSPISLTILAIISARARPKYKNNPVYKGTIGIDSEKFSIKSGSKEDYSRIIK  
ELSEIAKVNK"  
CDS complement(8399. .9261)  
/gene="B80676"  
/complement(8399. .9261)  
/gene="B80676"  
/note="similar to GB:LA2023 SP:P44755 PID:1004013  
PID:122302 PID:1204815 percent identity: 25.58;  
identified by sequence similarity; putative"  
/codon\_start=1

Query Match 100.0%; Score 19; DB 2; Length 13271;  
Best Local Similarity 100.0%; Pred. No. 1.3; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gatgattagcagagggtt 19  
|||||  
Db 141 GATGATTAGCAGAGGGTT 159

RESULT 2  
BB062900 1655 bp DNA BCT 15-JAN-1997  
LOCUS Borrelia burgdorferi flagellar filament outsheath protein (flaA)  
DEFINITION gene, complete cds, and chemotaxis histidine kinase (cheA) gene,  
partial cds.  
ACCESSION U62900  
NID 91575445  
VERSION U62900.1 GI:1575445  
KEYWORDS Lyme disease spirochete.  
SOURCE Borrelia burgdorferi  
ORGANISM Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
burgdorferi group.  
REFERENCE 1 (bases 1 to 1655)  
AUTHORS Ge-Y. and Charon N.W.  
TITLE An unexpected flaA homolog is present and expressed in Borrelia  
burgdorferi  
JOURNAL J. Bacteriol. 179 (2), 552-556 (1997)  
MEDLINE 97144545  
REFERENCE 2 (bases 1 to 1655)  
AUTHORS Ge-Y.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-1996) Yigong Ge, West Virginia University,  
Microbiology, HSCN, Morgantown, WV 26506, USA  
FEATURES  
source location/Qualifiers  
1. 1655  
/organism="Borrelia burgdorferi"  
/strain="212"  
/db\_xref="taxon:139"  
/clone="pw1 and pw2"

CDS <1. .395  
/function="unknown"  
/note="orfA"  
/codon\_start=3  
/transl\_table=11  
/protein\_id="AAC44769.1"  
/db\_xref="PID:g1575446"  
/db\_xref="GI:1575446"  
/translation="DNLFSEKRVFPFHSKGFVANLRVLRDEONLKDNLVYKDFLLG  
SNESFSGFLSDSRSLYSLKNGYVYVNSREFYDSFNNGDYNESDYVNFANSL  
IKTRMRYNPKIKIKIILIVEGCIKESD"  
473. .1498  
/gene="flaA"  
473. .1498  
/gene="flaA"  
/function="putative flagellar filament outsheath protein"  
/codon\_start=1  
/transl\_table=11  
/product="flaA protein"  
/protein\_id="AAC44770.1"  
/db\_xref="PID:g1575447"  
/db\_xref="GI:1575447"  
/translation="MKRKASILFPLISTVLAFOETDGLAEGSKRAEPGEVLVDFAEI  
ARDPSSSTRDLTNYVDYVYSGAGIVKPEDMVVDLGINNSVLLTPSARLQAVKRSV  
VAPAVVSESKRYAGDTILGVRLVLPFSYSSQSSAMIPPEKIPFYSCEGNOFGKLI  
DNIKTMKEIKVSYSLGVEIDLEVLFDNMGMYAVSMGTLKFKGNADLI MSPNIP  
NISRRIIKDVPNYPVPLASSKMRKARVSKSHSKNFFYVKDLRVLYDKLSVID  
SDIDSESVFVYETSGTESLRKLAHETFRVYLKLEKISIAEGSFONFVEKIESKP  
EESSPKN"  
1566. .1655  
/gene="cheA"  
1566. .1655  
/gene="cheA"  
/function="chemotaxis histidine kinase"  
/codon\_start=1  
/transl\_table=11  
/product="CheA protein"  
/protein\_id="AAC44771.1"  
/db\_xref="PID:g1575448"  
/db\_xref="GI:1575448"  
/translation="METLDLENELLGVFFEEAQNVLILEINI"

BASE COUNT 560 a 182 c 308 g 605 t  
ORIGIN

Query Match 100.0%; Score 19; DB 2; Length 1655;  
Best Local Similarity 100.0%; Pred. No. 1.2; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gatgattagcagagggtt 19  
|||||  
Db 539 GATGATTAGCAGAGGGTT 557

RESULT 3  
HSA255N20 175001 bp DNA HMG 11-JUN-1999  
LOCUS Homo sapiens chromosome 22 clone A255N20, WORKING DRAFT SEQUENCE,  
DEFINITION in unordered pieces.  
ACCESSION AL078632  
NID 95051845  
VERSION AL078632.1 GI:5051845  
KEYWORDS HMG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 175001)  
AUTHORS Alnscough, R.  
TITLE Direct Submission  
JOURNAL Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

## COMMENT

IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished: ba255n20 Contig\_ID: 00808 acc- Length: 3623 bp Unfinished: ba255n20 Contig\_ID: 00926 acc- Length: 1023 bp Unfinished: ba255n20 Contig\_ID: 00649 acc- Length: 1702 bp Unfinished: ba255n20 Contig\_ID: 00966 acc- Length: 1146 bp Unfinished: ba255n20 Contig\_ID: 00011 acc- Length: 9905 bp Unfinished: ba255n20 Contig\_ID: 00211 acc- Length: 3175 bp Unfinished: ba255n20 Contig\_ID: 01060 acc- Length: 1430 bp Unfinished: ba255n20 Contig\_ID: 01061 acc- Length: 1116 bp Unfinished: ba255n20 Contig\_ID: 00135 acc- Length: 1267 bp Unfinished: ba255n20 Contig\_ID: 01144 acc- Length: 1731 bp Unfinished: ba255n20 Contig\_ID: 00653 acc- Length: 16059 bp Unfinished: ba255n20 Contig\_ID: 00734 acc- Length: 3766 bp Unfinished: ba255n20 Contig\_ID: 00456 acc- Length: 4990 bp Unfinished: ba255n20 Contig\_ID: 00771 acc- Length: 7101 bp Unfinished: ba255n20 Contig\_ID: 00022 acc- Length: 2031 bp Unfinished: ba255n20 Contig\_ID: 00420 acc- Length: 1810 bp Unfinished: ba255n20 Contig\_ID: 00180 acc- Length: 9266 bp Unfinished: ba255n20 Contig\_ID: 01190 acc- Length: 3028 bp Unfinished: ba255n20 Contig\_ID: 00381 acc- Length: 2378 bp Unfinished: ba255n20 Contig\_ID: 01191 acc- Length: 2371 bp Unfinished: ba255n20 Contig\_ID: 00068 acc- Length: 4820 bp Unfinished: ba255n20 Contig\_ID: 01194 acc- Length: 3603 bp Unfinished: ba255n20 Contig\_ID: 00114 acc- Length: 1983 bp Unfinished: ba255n20 Contig\_ID: 000270 acc- Length: 7708 bp Unfinished: ba255n20 Contig\_ID: 00631 acc- Length: 2473 bp Unfinished: ba255n20 Contig\_ID: 01162 acc- Length: 3548 bp Unfinished: ba255n20 Contig\_ID: 01128 acc- Length: 1168 bp Unfinished: ba255n20 Contig\_ID: 00355 acc- Length: 5705 bp Unfinished: ba255n20 Contig\_ID: 00193 acc- Length: 1040 bp Unfinished: ba255n20 Contig\_ID: 00951 acc- Length: 3444 bp Unfinished: ba255n20 Contig\_ID: 00677 acc- Length: 2575 bp Unfinished: ba255n20 Contig\_ID: 00957 acc- Length: 2708 bp Unfinished: ba255n20 Contig\_ID: 01053 acc- Length: 4295 bp Unfinished: ba255n20 Contig\_ID: 00207 acc- Length: 1762 bp Unfinished: ba255n20 Contig\_ID: 01018 acc- Length: 1196 bp Unfinished: ba255n20 Contig\_ID: 00127 acc- Length: 1631 bp Unfinished: ba255n20 Contig\_ID: 00443 acc- Length: 1276 bp Unfinished: ba255n20 Contig\_ID: 00920 acc- Length: 4305 bp Unfinished: ba255n20 Contig\_ID: 00482 acc- Length: 4755 bp Unfinished: ba255n20 Contig\_ID: 00087 acc- Length: 1870 bp Unfinished: ba255n20 Contig\_ID: 00528 acc- Length: 1526 bp.

\* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

## FEATURES

## source

Location/Qualifiers  
 1..175001  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="22"  
 /clone="A255N20"

BASE COUNT 36468 a 34744 c 33549 g 37416 t 32824 others  
 ORIGIN

Query Match 89.5%; Score 17; DB 34; Length 175001;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gatgatgacagagg 17  
 ||||||||||||||||  
 Db 14964 GATGATTACAGAGG 14980

RESULT 4  
 HSB9F11/c

## LOCUS

HSBA9F11 188759 bp DNA HTG 10-JUN-1999  
 DEFINITION Homo sapiens chromosome 22 clone BA9F11, WORKING DRAFT SEQUENCE, in  
 unordered pieces.

## ACCESSION

95051328  
 AL050312.4 GI:5051328

## VERSION

HTG; HTGS; PHASE1.  
 human.

## KEYWORDS

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Euthera; Primates; Catarrhini; Homiidae; Homo.

## SOURCE

1 (bases 1 to 188759)  
 Laid, G.

## REFERENCE

Submitted (10-JUN-1999) Wellcome Trust Genome Campus, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquere@anger.ac.uk Clone requests: clones@anger.ac.uk

## AUTHORS

On Jun 11, 1999 this sequence version replaced gi:5042228.  
 IMPORTANT: This sequence is unfinished and does not necessarily  
 represent the correct sequence. Work on the sequence is in progress  
 and the release of this data is based on the understanding that the  
 sequence may change as work continues. The sequence may be  
 contaminated with foreign sequence from E.coli, yeast, vector,  
 phage etc. Order of segments is not known; 800 n's separate  
 segments. Unfinished: BA9F11 Contig\_ID: 02828 acc-AL050312  
 Length: 1018 bp Unfinished: BA9F11 Contig\_ID: 01303 acc-AL050312  
 Length: 1045 bp Unfinished: BA9F11 Contig\_ID: 01541 acc-AL050312  
 Length: 1100 bp Unfinished: BA9F11 Contig\_ID: 02365 acc-AL050312  
 Length: 48381 bp Unfinished: BA9F11 Contig\_ID: 02368 acc-AL050312  
 Length: 1194 bp Unfinished: BA9F11 Contig\_ID: 01478 acc-AL050312  
 Length: 4316 bp Unfinished: BA9F11 Contig\_ID: 00710 acc-AL050312  
 Length: 1048 bp Unfinished: BA9F11 Contig\_ID: 03060 acc-AL050312  
 Length: 1050 bp Unfinished: BA9F11 Contig\_ID: 02133 acc-AL050312  
 Length: 1168 bp Unfinished: BA9F11 Contig\_ID: 02252 acc-AL050312  
 Length: 114113 bp Unfinished: BA9F11 Contig\_ID: 01923  
 acc-AL050312 Length: 2401 bp Unfinished: BA9F11 Contig\_ID: 00557  
 acc-AL050312 Length: 1224 bp Unfinished: BA9F11 Contig\_ID: 00403  
 acc-AL050312 Length: 1101 bp.

\* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

## FEATURES

## source

Location/Qualifiers  
 1..188759  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="22"  
 /clone="BA9F11"

BASE COUNT 44912 a 41540 c 44321 g 48377 t 9609 others  
 ORIGIN

Query Match 89.5%; Score 17; DB 34; Length 188759;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gatgatgacagagg 17  
 ||||||||||||||||  
 Db 19018 GATGATTACAGAGG 19002

## RESULT 5

## HS477H23

## LOCUS

## DEFINITION

## ACCESSION

## NID

## VERSION

## KEYWORDS

## HTG.

## SOURCE

## ORGANISM

HS477H23 114600 bp DNA PRI 01-MAR-1999  
 DEFINITION Human DNA sequence from clone 477H23 on chromosome 22q12.1-12.2.  
 Contains parts of one or two novel genes. Contains ESTs and GSSs,  
 complete sequence.

## ACCESSION

## NID

## VERSION

## KEYWORDS

## HTG.

## SOURCE

## ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 114600)  
Clark, G.  
Direct Submission  
Submitted (28-JAN-1999) Sanger Centre, Hinxton, Cambridgeshire

[illegible]

Query Match	86.3%	Score 16.4;	DB 10;	Length 114600;
Best Local Similarity	94.4%;	Pred. No. 38;		
Matches 17; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

```

QY      1 gatgattacagaggt 18
        ||||| |||||
Db 17925 GATGAGTAGCAGAGGCT 17942

```

RESULT	6
LOCUS	HSB107C2
DEFINITION	Human genomic DNA sequence from cosmid B107C2 on chromosome 17-ANG-1998
ACCESSION	AJ008724
NID	92760024
VERSION	AL008724.1 GI:2760024
KEYWORDS	22q11.2-12.1.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE 1 (bases 1 to 44665)

## AUTHORS

**TITLE** Direct Submission

JOURNAL

Submitted (29-DEC-1997) Chromosome 22 Project Group  
(<http://www.sanger.ac.uk/HGP/Chr22/>) Sanger Centre, Hinxton  
Cambridgeshire, CB10 1SA, UK. E-mail enquires:

**COMMENT**

hungry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
On Jan 8, 1998 this sequence version replaced gi:2598506.  
**IMPORTANT:** This sequence is the entire insert of clone B107C2. This  
sequence has been finished according to sequence map criteria as  
follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The true left end of clone B107C2 is at 1 in this sequence. The  
true right end of clone B107C2 is at 4465.  
B107C2 is from the human chromosome 22-specific cosmid library  
(SC22CB) constructed at the Sanger Centre by Mark Ross and Coriella  
Langford. VECTOR: Lawrist 16.  
This sequence was generated from part of bacterial clone contigs of  
human chromosome 22 constructed by the Sanger Centre chromosome 22  
mapping group.  
Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr22/>.

**FEATURES**  
**source**

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2-12.1"
/clone="B107C2"
/clone_1fb="SC22c8"
1233. 1366
repeat_region
/note="FLAM_A repeat: matches 1. .132 of consensus"
2113. 2146
repeat_region
/note="17 copies of 2 mer 85 & conserved"
3417. .3675
repeat_region
/note="PLMA2 repeat: matches 1055. .782 of consensus"
4335. .5562
prim_transcript
/note="match: ESTs AA393708 RS8963 W26684 W76562 W72558
D60118 D60119 F11418 W25596 D80911 F01844 H61750 D60539
AA026630 AA010068 AA340208 H61749 H40968 D81868 AA116109
R614231 T35711 AA400907 T64146 R59018 F11011 R37806 T64274;
poly-A tail on several clones at position 4335"
4343. 4585
misc_feature
/note="match: STS G05242"
5521. .5814
repeat_region
/note="AluX repeat: matches 1. .296 of consensus"
5599. .6606
prim_transcript
/note="match: ESTs D30845 R61460 F05620 N55842 F05577
N25729"
N25729
prim_transcript
/note="complement(7040. .7546)
paired with EST N36440; paired with EST N25729 matching
this clone"
7546. .7573
repeat_region
/note="14 copies of 2 mer 100 & conserved"
7592. .8329
prim_transcript
/note="match: ESTs AA496599 AA129336 R55270 H10979"
8356. .8974
prim_transcript
/note="match: ESTs AA443013 R55269 H10978 AA359630; R55269
paired with EST R55270 matching this clone"
10868. .11267
prim_transcript
/note="match: EST AA354647"
11605. .12279
prim_transcript
/note="match: ESTs R19881 R13060"
12404. .12747
repeat_region
/note="match: ESTs H22044 T85019 T72724 H45918"
13756. .14026
repeat_region
/note="THEIC repeat: matches 1. .371 of consensus"
14238. 14374
repeat_region
/note="Aluud repeat: matches 1. .294 of consensus"
14391. .14894
repeat_region
/note="Aluud repeat: matches 1. .133 of consensus;
incomplete repeat"
15223. 15357
repeat_region
/note="AluSp repeat: matches 303. .2 of consensus"
15223. 15357
repeat_region
/note="MIR2 repeat: matches 8. .143 of consensus"

```

```
repeat_region 15857..16056
/note="Aluub repeat: matches 200. .1 of consensus;
incomplete repeat"
repeat_region 20476..20590
/note="Aluub repeat: matches 6. .120 of consensus;
incomplete repeat"
repeat_region 21935..22082
/note="MIR repeat: matches 204. .69 of consensus"
repeat_region 24416..24502
/note="MIR repeat: matches 158. .70 of consensus"
repeat_region 29030..29314
/note="Aluux repeat: matches 1. .301 of consensus"
repeat_region 32536..32689
/note="MIR repeat: matches 96. .257 of consensus"
repeat_region 36055..36128
/note="MIR repeat: matches 67. .146 of consensus"
repeat_region 37194..37286
/note="MIR2 repeat: matches 139. .37 of consensus"
repeat_region 37337..37494
/note="MIR2 repeat: matches 1. .160 of consensus"
repeat_region 37619..37904
/note="Aluub repeat: matches 2. .287 of consensus"
repeat_region 37911..37974
/note="MIR2 repeat: matches 278. .345 of consensus"
repeat_region 38826..39115
/note="Aluub repeat: matches 290. .1 of consensus"
repeat_region 39163..39304
/note="MIR repeat: matches 154. .14 of consensus"
repeat_region 40044..40135
/note="2 copies of 46 mer 99 & conserved"
repeat_region 41434..41524
/note="MIR2 repeat: matches 146. .55 of consensus"
repeat_region 41750..42029
/note="Aluub repeat: matches 281. .5 of consensus;
incomplete repeat"
repeat_region 44553..44618
/note="33 copies of 2 mer 94 & conserved"
BASE COUNT 11149 a 11147 c 11427 g 10942 t
ORIGIN
```

```
Query Match 86.3%; Score 16.4; DB 10; Length 4465;
Best Local Similarity 94.4%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 gatgattagcagaggt 18
||||| |||||||
Db 5202 GATGAGTACGAGGCT 5219
```

```
RESULT 7
HS353E16 226295 bp DNA HTG 11-JUN-1999
LOCUS Homo sapiens chromosome 22 clone 353E16, WORKING DRAFT SEQUENCE, 1n
DEFINITION unoriented pieces.
ACCESSION AL031591
NID 95051837
VERSION AL031591.11 GI:5051837
KEYWORDS HTG; HTGS_PHA5E1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 226295)
REFERENCE 1
AUTHORS Clark,G.
JOURNAL Direct Submision
Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Jun 12, 1999 this sequence version replaced gi:5042216.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
```

sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known: 800 n's separate segments. Unfinished: dJ353E16 Contig\_ID: 03404 acc-AL031591 length: 73524 bp unfinished: dJ353E16 Contig\_ID: 03127 acc-AL031591 Length: 151971 bp.

\* NOTE: This is a 'working draft' sequence.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

#### FEATURES

##### source

```
1..226295
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="353E16"
```

```
BASE COUNT 66949 a 50625 c 49283 g 58638 t 800 others
ORIGIN
```

```
Query Match 86.3%; Score 16.4; DB 34; Length 226295;
Best Local Similarity 94.4%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 gatgattagcagaggt 18
||||| |||||||
Db 17950 GATGAGTACGAGGCT 17933
```

```
RESULT 8
HS692C8 125698 bp DNA PRI 07-APR-1999
LOCUS Human DNA sequence from clone 692C8 on chromosome 20p11.22-12.2
DEFINITION Contains STRs and GSSs, complete sequence.
ACCESSION AL034561
NID 94464247
VERSION AL034561.4 GI:4464247
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 125698)
REFERENCE 1
AUTHORS Matthews,L.
JOURNAL Direct Submision
Submitted (07-APR-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Mar 22, 1999 this sequence version replaced gi:4455600.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence is the entire insert of clone 692C8. This sequence
has been finished according to sequence map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
692C8 is from the library RPC14 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: pCYRAC.
```

#### FEATURES

##### source

```
1..125698
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```
repeat_region /chromosome="20"
/clone="692C8"
/map="p11.22-12.2"
/clone.lib="RPC14"
1.1506
repeat_region /note="L1P repeat: matches 617. .2120 of consensus"
1501. .1586
repeat_region /note="L1P repeat: matches 13. .98 of consensus"
1644. .1844
repeat_region /note="L2 repeat: matches 2527. .2721 of consensus"
1863. .2193
repeat_region /note="L1R16C repeat: matches 2. .353 of consensus"
3271. .3738
repeat_region /note="L1M4 repeat: matches 7466. .7968 of consensus"
4629. .4809
repeat_region /note="L2 repeat: matches 2353. .2528 of consensus"
4843. .4956
repeat_region /note="L1R1H repeat: matches 434. .547 of consensus"
5762. .5870
repeat_region /note="MER5A repeat: matches 9. .109 of consensus"
6398. .6748
repeat_region /note="THE1A repeat: matches 5. .354 of consensus"
6776. .7080
repeat_region /note="Alusg repeat: matches 1. .302 of consensus"
7267. .7403
repeat_region /note="MIR repeat: matches 12. .146 of consensus"
7579. .7847
repeat_region /note="L1R1H repeat: matches 285. .547 of consensus"
8026. .8655
repeat_region /note="L1M3A repeat: matches 5279. .5899 of consensus"
8656. .9173
repeat_region /note="L1P8 repeat: matches 5642. .6161 of consensus"
9174. .9373
repeat_region /note="L1M3A repeat: matches 5082. .5279 of consensus"
9402. .9537
repeat_region /note="L1P83 repeat: matches 5999. .6148 of consensus"
10921. .11094
repeat_region /note="MER5A repeat: matches 10. .189 of consensus"
11322. .11389
repeat_region /note="L2 repeat: matches 2674. .2742 of consensus"
11391. .11752
repeat_region /note="MER39 repeat: matches 8. .380 of consensus"
11797. .11912
repeat_region /note="MER39b repeat: matches 429. .550 of consensus"
11943. .12067
repeat_region /note="L2 repeat: matches 2543. .2670 of consensus"
14435. .14490
repeat_region /note="28 copies 2 mer aa 73% conserved"
15098. .15113
repeat_region /note="L1P82 repeat: matches 5729. .6152 of consensus"
15514. .15575
repeat_region /note="31 copies 2 mer ta 77% conserved"
16460. .16868
repeat_region /note="MER7B repeat: matches 1. .418 of consensus"
16883. .17029
repeat_region /note="MIR repeat: matches 3. .154 of consensus"
17244. .17532
repeat_region /note="L1R16C repeat: matches 1. .283 of consensus"
17666. .18101
repeat_region /note="MSTC repeat: matches 1. .401 of consensus"
19055. .19503
repeat_region /note="MIR repeat: matches 9. .262 of consensus"
20187. .20340
repeat_region /note="L1R1H repeat: matches 252. .397 of consensus"
21586. .21878
repeat_region /note="L2 repeat: matches 2425. .2745 of consensus"
22887. .22983
repeat_region /note="MIR repeat: matches 48. .145 of consensus"
24076. .24233
repeat_region /note="MER5B repeat: matches 2. .178 of consensus"
24343. .24580
repeat_region /note="L1M5A repeat: matches 6042. .6285 of consensus"
24791. .24890

repeat_region /note="MIR repeat: matches 21. .121 of consensus"
26211. .26639
repeat_region /note="L1R16A repeat: matches 16. .449 of consensus"
27314. .27378
repeat_region /note="L2 repeat: matches 2685. .2750 of consensus"
27391. .27432
repeat_region /note="21 copies 2 mer tt 83% conserved"
27774. .27814
repeat_region /note="L2 repeat: matches 2632. .2674 of consensus"
27871. .28175
repeat_region /note="AluJo repeat: matches 1. .300 of consensus"
28323. .30292
repeat_region /note="L1P82 repeat: matches 4177. .6146 of consensus"
30322. .30623
repeat_region /note="Alusg repeat: matches 4. .297 of consensus"
31832. .32141
repeat_region /note="AlusX repeat: matches 1. .307 of consensus"
33639. .34021
repeat_region /note="L1R1H repeat: matches 29. .494 of consensus"
34345. .34663
repeat_region /note="L1M3C repeat: matches 10. .336 of consensus"
35259. .35554
repeat_region /note="AluJb repeat: matches 2. .308 of consensus"
35356. .35720
repeat_region /note="L1M3C repeat: matches 1470. .1634 of consensus"
35848. .39100
repeat_region /note="L1M1 repeat: matches 1554. .4497 of consensus"
39103. .39247
repeat_region /note="AluJo repeat: matches 1. .210 of consensus"
39248. .39338
repeat_region /note="AluJo repeat: matches 118. .209 of consensus"
39345. .39978
repeat_region /note="L1M410 repeat: matches 5677. .6316 of consensus"
41105. .41405
repeat_region /note="L1P816 repeat: matches 5850. .6150 of consensus"
complement(<41293. .41574)
/note=match: GSS A0098779
41865. .41922
repeat_region /note="MIR repeat: matches 91. .149 of consensus"
42854. .43166
repeat_region /note="AluY repeat: matches 1. .311 of consensus"
43891. .43903
repeat_region /note="L1 repeat: matches 3000. .3011 of consensus"
43904. .44683
repeat_region /note="L1M4A repeat: matches 5516. .6295 of consensus"
44684. .45092
repeat_region /note="L1 repeat: matches 2583. .3000 of consensus"
45069. .46984
repeat_region /note="L1 repeat: matches 91. .2482 of consensus"
46916. .48259
repeat_region /note="L1M2 repeat: matches -674. .654 of consensus"
complement(49576. .49922)
/note=match: 294578 chromosome 20 HindIII fragment"
49813. .50065
repeat_region /note="L1M2C repeat: matches 1728. .1978 of consensus"
50165. .50530
repeat_region /note="THE1C repeat: matches 1. .371 of consensus"
50799. .51248
repeat_region /note="L1M2 repeat: matches 3855. .4316 of consensus"
51298. .51546
repeat_region /note="L1M2 repeat: matches 4323. .4567 of consensus"
51547. .51773
repeat_region /note="MER58A repeat: matches 2. .224 of consensus"
51774. .52319
repeat_region /note="L1M2 repeat: matches 4567. .5129 of consensus"
52461. .52618
repeat_region /note="L1M2 repeat: matches 5281. .5436 of consensus"
52921. .53079
repeat_region /note="L1M4 repeat: matches 5761. .5924 of consensus"
53763. .53936
repeat_region /note="MER5B repeat: matches 13. .117 of consensus"
54968. .55107
repeat_region /note="L1R16C repeat: matches 247. .387 of consensus"
```

repeat\_region 55123..55260  
/note="LMB7 repeat: matches 6032..6171 of consensus"  
repeat\_region 55299..55460  
/note="LMB7 repeat: matches 70..235 of consensus"  
repeat\_region 55709..55812  
/note="MER34 repeat: matches 6..106 of consensus"  
repeat\_region 55771..55962  
/note="LMB7 repeat: matches 29..227 of consensus"  
repeat\_region 56116..56198  
/note="MER34 repeat: matches 459..540 of consensus"  
repeat\_region 56696..56839  
/note="72 copies 2 mer aa 58% conserved"

Query Match 84.2%; Score 16; DB 10; Length 125698;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 ggattagcagaggtt 19  
|||||  
Db 5569 GGATTAGCAGAGGTT 5584

RESULT 9  
LOCUS PIP404BCN 3105 bp DNA BCT 07-MAY-1996  
DEFINITION Plasmid PIP404 (from Clostridium perfringens) bacteriocin (bcn5),  
complete cds.  
ACCESSION M14481  
M14481  
VERSION M14481.1 GI:150736  
KEYWORDS bacteriocin; bacteriocin BCNF.  
SOURCE Plasmid PIP404 (clone: pT915.) DNA.  
ORGANISM plasmid PIP404

REFERENCE  
AUTHORS 1 (bases 1 to 3105)  
TITLE Garnier, T. and Cole, S. T.  
Characterization of a bacteriocinogenic plasmid from Clostridium  
perfringens and molecular genetic analysis of the  
bacteriocin-encoding gene  
J. Bacteriol. 168 (3), 1189-1196 (1986)  
87057020  
JOURNAL  
MEDLINE  
COMMENT Draft entry and computer-readable sequence of [1] kindly provided  
by S. T. Cole, 10-FEB-1987.  
A region of dyad symmetry (21 bp inverted repeats with 5 bp in  
between) is found at positions 2973-3019.

FEATURES  
Source location/Qualifiers

1..3105  
/organism="Plasmid PIP404"  
/plasmid="Plasmid PIP404"  
/specific\_host="CPN50"  
/db\_xref="taxon:2573"  
/clone="pT915."  
301..2973  
/gene="bcn5"  
301..2973  
/gene="bcn5"  
/codon\_start=1  
/transl\_table=11  
/product="bacteriocin"  
/protein\_id="AA08248.1"  
/db\_xref="PID:g150737"  
/db\_xref="GI:150737"  
/translation="MANNIIPNVSGLVSTPFPNNAVVRGDELRLRVNDGQIPG  
RTVSDGEITVLPSISNEKNTVLVQPSGSGRGQYVNAATSIKKYHDYSWNSSTPE  
PVYDEKTOIGLIDPREKAVLVKTVGKATYVATDTGKGLTSGLVHTEGSSSTIG  
GSFNGVAPGEVPEGFTYENNAEYVDELTLRDANGVLIGRSVSGDKITVLDVGT  
KQALVQYPAQDVROGYVNTATNLIRYFQYVSWHNSSTSEVYLDENGHLSINPTE  
AATLLEKNGMKHVVYDTNKGPNKSGYVEGAATRVDPISITNAOKIYGIIG  
RGRELAAVKGNGSNLSVFCALHIGEDNNAADIELTRIGNGLIIEHONAGNNMEL  
YIIPVAPDGLSEGTNNGRCYVAVDNCNDPFGFSGVPRYHSGSEPLSVSE  
SKSLHDFIOGKNTSGEMCYVDHGHGEGALIGPELGEFRNDFGQSGQSGIGDNRG  
PMIGMASTGAKKALIELPGSTKSHSVNGRKYQKTIINAVTNLIGSGSSSSGSSSF  
SDVSEATGGEIVNQSLNVEGAGLYTNSIGLRQGNKVNIAKNGDMYKIKYXGSEY

GIYVNSGYIIILKNNISVKLEDMQEDCIKGCWPIKEKLEIYWDSTRLYKSIENDISQ  
AARKSLINVINPLNFSVSEMIACQIVNNNETSPFRBEMSKSNPNIVYKSLSN  
GOIIVLDRIINIKPEKTKTIPRAAGAFKDTIKFEFFKIDGMFTAIAGALSIGDL  
SVRQSGELKSNNDIKALAAAYIVNGVETMPCAFIAGFLAQIAIEFPVIAVAGAI  
VSAIAPALGIFPLDNEKREKILMNSKGLIDYLF"  
BASE COUNT 1248 a 326 c 596 g 935 t  
ORIGIN 225 bp upstream of HindIII site.

Query Match 83.2%; Score 15.8; DB 1; Length 3105;  
Best Local Similarity 89.5%; Pred. No. 82;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gatgattagcagaggtt 19  
|||||  
Db 1501 GATGATTAGCAGAGGTT 1519

RESULT 10  
LOCUS PIP404CG 10206 bp DNA circular BCT 07-MAY-1996  
DEFINITION Plasmid PIP404 (from Clostridium perfringens), complete genome.  
ACCESSION M32882 J03309 J03310  
M32882 J03309 J03310  
VERSION M32882.1 GI:150738  
KEYWORDS bacteriocin; bcn gene; complete genome; uv1AB gene; uv1B gene.  
SOURCE Plasmid PIP404 DNA.  
ORGANISM plasmid PIP404

REFERENCE  
AUTHORS 1 (bases 775 to 1224; 7031 to 8430)  
TITLE Garnier, T. and Cole, S. T.  
Studies of UV-inducible promoters from Clostridium perfringens in  
vivo and in vitro  
Mol. Microbiol. 2 (5), 607-614 (1988)  
89039249  
JOURNAL  
MEDLINE  
COMMENT Draft entry and computer-readable sequence for [1], [2] kindly  
submitted by T. Garnier, 15-DEC-1988.  
location/Qualifiers

FEATURES  
Source location/Qualifiers

1..10206  
/organism="Plasmid PIP404"  
/plasmid="Plasmid PIP404"  
/specific\_host="Clostridium perfringens"  
/db\_xref="taxon:2573"  
855..901  
/note="putative"  
complement(900..1496)  
/standard\_name="ORF 4"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA08250.1"  
/db\_xref="PID:g455315"  
/db\_xref="GI:455315"  
/translation="MIIKSDTKDLISILITFLVLCNLIVFSSNKAVENQSIPL  
MMLLMNFIILISLIVLFLKLVKPPALNLGLIFSSSLODILISIMNFQILS  
FSKFIYFIIFIIYILFLSFYLAIRKRLKNNKNTSKPKNTIFIIAALIGIVFSKI  
NENIPQYIISFLMYVSCFTGFGFYHIYIIVSNKN"  
1091..1144  
/note="putative"  
1832..3052  
/standard\_name="ORF 5"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA08251.1"  
/db\_xref="PID:g294239"  
/db\_xref="GI:294239"  
/translation="MRNALNKNKNYIDNHSKQWITRSVIDKGYSQWHYKVAL  
KIDMSDENIYITLNTFYKPCRLLENIKELNTLFDLDYKTKGTQDOVLMDEKNYF

CDS  
3809..4309  
/standard\_name="ORF 6"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA98252.1"  
/db\_xref="pid:g150742"  
/db\_xref="gi:150742"  
/translation="MLRGVDYAEKELKYSKTAIYKRLKEFRKRVYKQKQSMIDEE  
LFNLKIDSLKYNVEYKDYIEESKDEKSEIAMDREGSLNKLIDTLIAOLEEKD  
KOAIELHLIENNOVLAKKEQETKINILEFEHFKEVONKLSSIEKKNQREKKSFF  
KNFEKK"  
4347..4514  
/standard\_name="ORF 7"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA98253.1"  
/db\_xref="pid:g150743"  
/db\_xref="gi:150743"  
/translation="MKNIDSIKRLIFSSICRAIAGILGLVDKNVLLGGAFTILMVSNI  
ILISEKKRLK"  
4919..5488  
/standard\_name="ORF 8"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA98254.1"  
/db\_xref="pid:g150744"  
/db\_xref="gi:150744"  
/translation="MLVGARYSTEEOSLNROIIDLVDYGVDRNIYQEKISGMKPNR  
BOLDKMIDELQGDVITITDLRISRSRTKLNIIDRIKAGASIKSIKIDMTDSSD  
NPYNSFLITVAGSLQSLERDLISQTRKGLSAKARGNGRPRKRNDAQTVGILIR  
EGKIVDIAVKGSLSRATVRYRLNDLKLK"  
5514..5870  
/standard\_name="ORF 9"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA98255.1"  
/db\_xref="pid:g150745"  
/db\_xref="gi:150745"  
/translation="MIMSKLSSINYYINKRIWGEHLKENILLINOYTIEDAFILEDG  
IKYLDKRTVEYIDLSEEDMKIEEAFIERLEKKRKVNKDKENFKNHMIMITEYLENE  
KSKEKSNVIELKNYRK"  
complement(5962..6990)  
/standard\_name="ORF 10"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA98256.1"  
/db\_xref="pid:g150746"  
/db\_xref="gi:150746"  
/translation="WKIARGGHNF LANGAGCLIDETIEDRKVYKAIKNIENFEV  
LDVTPGDDIDINTDLKLVGDKANNFNADLFISHPKCYDKDFGIGTGTWCEKGRKA  
EIIAQNIVDTISETGSLKNRGVTKNAKYLENTIPAVIYEVCESEKAVVDIYREK  
GSDILGYLIAGKICSVNKEISSDLPVYNLENTTSONNNLEKTNATAKVALDPDNP  
SNYKDIETIENERIKTILAEVCDLKFLPATYMODSINKESSPTWVSQKOTVAVDT  
NATVINVITELDARTPPSPDSNRMGIVANQERLVEHAKIENNATLATYTLASEGYKAMF  
IAEYTKLD  
-35\_signal  
7103..7108  
misc\_feature  
7108..7158  
/note="region of dyad symmetry"  
7121..7148  
/note="region of dyad symmetry; putative"  
7126..7133  
7148..>7858  
/gene="uv1AB"  
/note="alternative transcript"  
7148..7858  
/gene="uv1AB"  
7176..7181  
-35\_signal

-10\_signal  
7199..7204  
/gene="uv1AB"  
7218..>7858  
/gene="uv1AB"  
/note="alternative transcript"  
7286..7291  
/gene="uv1AB"  
7301..7858  
/gene="uv1AB"  
/codon\_start=1  
/transl\_table=11  
/product="bacteriocin"  
/protein\_id="AAA98257.1"  
/db\_xref="pid:g150747"  
/db\_xref="gi:150747"  
/translation="MSLEYKNIVYQNGDKKAIYIINPEIILNKYKSLKEIHFN  
SYIDENKQDLYSLINIVNKIPIDNQFENEGCLVATYIKSILNSKDMYINKNIR  
VFIESQSLSSWVEFKDKPLVKYIESNIEIDMLKCLEKEQKVIKRYFLNDKSEVEIA  
EIMGTSRQWIRIRIKTALKIKENI"  
7925..7930  
7939..8133  
/gene="uv1B"  
7939..8133  
/gene="uv1B"  
/codon\_start=1  
/transl\_table=11  
/product="bacteriocin"  
/protein\_id="AAA98258.1"  
/db\_xref="pid:g150748"  
/db\_xref="gi:150748"  
/translation="MDESLFKMATOGAFAILFSYLLFYVLKENSREKREYONIEEL  
TELPKIKEDVEDIEKRLNK"  
8140..8167  
/note="putative"  
8217..10206  
/note="bcn mRNA (major alt.)"  
join(8217..10206,1..855)  
/note="major alternative transcript"  
8264..8269  
8287..8294  
8288..10206  
/note="minor alternative transcript"  
join(8288..10206,1..855)  
/note="minor alternative transcript"  
join((8298..8300)..10206,1..855)  
/gene="bcn"  
/note="alternative transcript"  
join((8298..8300)..10206,1..855)  
/gene="bcn"  
8369..8381  
/gene="bcn"  
join(8389..10206,1..855)  
/gene="bcn"  
/codon\_start=1  
/transl\_table=11  
/product="bacteriocin"  
/protein\_id="AAA98249.1"  
/db\_xref="pid:g150739"  
/db\_xref="gi:150739"  
/translation="MANNIIPNVSSGDLVGSPTPEPPNAVVRGDFLYLRDVGNOIPG  
RTVSDGDEITVLFISNEKNIVLYQPTSSGRQGYVNASIIKRYKDYSVNGSTGE  
PVYDEKTOIGTLDPREBAVLYLVKVGDMNAADIELETRIGNGLEIHFONAGTVNGSL  
GSEFVAGVEYVPGGFTYENNAEYVGGELVLRONGNLIGRSVSGDKITVLDVGYT  
KQLALVYPRAGDVYRQGYTNATVNLIRFPNDYSMHNSSTSEVLDENGCHIGSLNPE  
AATLYERKNKRAHVYDITNGPNTKSGVYIEGAAATRVDPYPSITNAOKIYVIGISG  
RGRBLAAYKVNGSNSLVFVCAIHGMDNMAADIELETRIGNGLEIHFONAGTVNGSL  
YIIRVANDGISSEGTINNGPGRCTIYGAVDCNRDPLGFSFGVPRYHSGSEPLSVSE  
SKSLHDFLOGKYNRTSGEMCVYIDHGEAGAIQNPETGEYFRNOFGFGSGSGDNG  
PMIGMARSIKAKALIELPGSTRKSHSVVNGRYOKTIIINVTNLIGSGSGSGSGSF  
SDVSEATGEYINQSTLNRREGGLTNSIGOLRQGNKNIYAKNDWIKRIKGESEY  
GYVNSGYIILIKNNTSVKLEDMQEDCIKFGNGPTKRYLEYMDSTRLYSIENDISQ  
AIKNKSLINVINPLNFSVSEMIACTQIVFNNETTSFPRDEWYSKSNPNFIVKYKKLNS





```

Db      5904 GATGATAGCATAGCGCTT 5886
||||| ||| |||||
RESULT 12
LOCUS   CSU23186/c          DNA          VRT          15-OCT-1996
DEFINITION Crotalus scutulatus PLA2-like pseudogene (psi-Mtx).
ACCESSION U23186
NID      91001869
VERSION  U23186.1   GI:1001969
KEYWORDS
SOURCE   Mojave rattlesnake.
ORGANISM Crotalus scutulatus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Lepidosauria;
          Squamata; Scleroglossa; Serpentes; Colubroidae; Viperidae;
          Crotalinae; Crotalus.
REFERENCE 1 (bases 1 to 2418)
AUTHORS   John T.R., Smith, J.J. and Kaiser, I.I.
TITLE      A phospholipase A2-like pseudogene retaining the highly conserved
           introns of Mojave toxin and other snake venom group II PLA2s, but
           having different exons
JOURNAL    DNA Cell Biol. 15 (8), 661-668 (1996)
MEDLINE    96365391
REFERENCE 2 (bases 1 to 2418)
AUTHORS   John T.R. and Kaiser, I.I.
TITLE      Direct Submission
JOURNAL    Submitted (21-MAR-1995) Ted R. John, Molecular Biology, Univ. of
           Wyoming, Box 3944 University Station, Laramie, WY 82071-3944, USA
FEATURES
   source
       1..2418
           /organism="Crotalus scutulatus"
           /sub_species="scutulatus"
           /db_xref="taxon:8737"
       525..530
           join(557..573,648..798,966..1098,1574..1674,1934..215
           /gene="psi-Mtx"
           join(557..573,648..798,966..1098,1574..1674,1934..215
           /gene="psi-Mtx"
           join(<974..1098,1574..1674,1934..2076)
           /gene="psi-Mtx"
           /note="PLA2-like protein; This protein probably does not
           exist in reality, but only as an ORF on a pseudogene"
           /codon_start=1
           /pseudo
       2133..2138
           /gene="psi-Mtx"
BASE COUNT  605 a      660 c      624 g      529 t
ORIGIN
polyA_signal
2133..2138
/gene="psi-Mtx"
polyA_signal
2133..2138
/gene="psi-Mtx"
Query Match      83.2%   Score 15.8;   DB 4;   Length 2418;
Best Local Similarity 89.5%;   Pred. No. 82;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1 gatgattacagaggggtt 19
||||| ||| |||||
Db      1890 GATGATAGCTAGCGGCTT 1872

```

MID 92281936 GI:2281936  
 VERSION 295351.1  
 KEYWORDS HTG; Brain Protein E46-like; CPG island; FBIN1; FIBULIN-1 ISOFORM B

PRECUSOR.

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 127583)  
AUTHORS Connor, R.  
JOURNAL Direct Submission  
TITLE Submitted (03-JUL-1997) E-mail enquiries: humquery@sanger.ac.uk  
COMMENT On Jul 28, 1997 this sequence version replaced g1:2094794.  
IMPORTANT: This sequence is the entire insert of clone 941F9. This  
sequence was generated from part of bacterial clone contigs of  
human chromosome 22, constructed by the Sanger Centre chromosome 22  
mapping group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr22/  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The true left end of clone 941F9 is at 1 in this sequence. The true  
left end of clone N38E2 is at 9038.  
The true right end of clone N38E2 is at 54042.  
The true left end of clone 398C22 is at 87881.  
The true right end of clone 941F9 is at 127583.  
941F9 is from the human BAC library described in U-D. Kim et al.  
(1996) Genomics 34, 213-218.  
VECTOR: pBelobAC11.

FEATURES  
Source  
Location/Qualifiers  
1..127583  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="22"  
/map="q11.2-qter"  
/clone="941F9"  
repeat\_region  
/note="16 copies of 2 mer 84 & conserved"  
2424..2647  
/note="7 copies of 32 mer 94 & conserved"  
2664..2714  
/note="MLR2A repeat: matches 453. .399 of consensus"  
2764..3168  
/note="MLR2B repeat: matches 409. .1 of consensus"  
3169..3420  
/note="MLR2 internal repeat: matches 5669. .5415 of  
consensus"  
3451..3752  
/note="AluX repeat: matches 302. .1 of consensus"  
3753..9137  
/note="MLR2 internal repeat: matches 5383. .2 of consensus"  
9139..9576  
/note="MLR2 repeat: matches 444. .1 of consensus"  
9520..9844  
/note="match: 5' ESTS C16630, C16762 - placenta"  
9777..9850  
/note="MIR repeat: matches 73. .141 of consensus"  
10365..10818  
/note="match: multiple ESTs; match: C18390 H02670 R23566  
R23317; match: R68138 R24691 R78762 T39562; match: R76273  
R62365 T49888 R68191; match: F16897 C02271 R23566; similar  
to FIBULIN-1 ISOFORM C PRECURSOR"  
10368..10691  
/note="match: STS R76273"  
11113..11148  
/note="MIR repeat: matches 144. .109 of consensus"  
11526..11571  
/note="MIR repeat: matches 120. .75 of consensus"  
12134..12269  
/note="MIR2 repeat: matches 4. .146 of consensus"  
12336..12815  
/note="match: multiple ESTs; match: R33281 H02060 R78704;  
similar to FIBULIN-1 ISOFORM B PRECURSOR"

misc\_feature  
12449..12848  
/note="match: STS G11168"  
13033..13136  
/note="MIR2 repeat: matches 21. .132 of consensus"  
13151..13198  
/note="12 copies of 4 mer 88 & conserved"  
13237..13335  
/note="AluX repeat: matches 2. .300 of consensus"  
14363..14606  
/note="MIR repeat: matches 259. .2 of consensus"  
14790..15080  
/note="AluX repeat: matches 1. .301 of consensus"  
15722..15855  
/note="MLTIC repeat: matches 442. .311 of consensus"  
15856..16026  
/note="3 copies of 57 mer 91 & conserved"  
16027..16197  
/note="3 copies of 57 mer 86 & conserved"  
16198..16311  
/note="2 copies of 57 mer 91 & conserved"  
16328..16624  
/note="MLTIC repeat: matches 308. .1 of consensus"  
16636..16814  
/note="MLTIB repeat: matches 1. .180 of consensus"  
16815..17114  
/note="AluX repeat: matches 1. .300 of consensus"  
17129..17326  
/note="MLTIB repeat: matches 164. .358 of consensus"  
17515..17795  
/note="AluX repeat: matches 22. .302 of consensus;  
incomplete repeat"  
18303..18587  
/note="AluX repeat: matches 1. .300 of consensus"  
18687..18732  
/note="LI repeat: matches 1784. .1885 of consensus"  
19069..19294  
/note="MER42C repeat: matches 45. .291 of consensus"  
19305..19603  
/note="AluX repeat: matches 1. .300 of consensus"  
19636..19885  
/note="MER42C repeat: matches 296. .549 of consensus"  
20401..20699  
/note="AluX repeat: matches 1. .302 of consensus"  
20850..21156  
/note="AluX repeat: matches 1. .300 of consensus"  
21517..21644  
/note="MIR2 repeat: matches 1. .145 of consensus"  
21967..47902  
/gene="BR941F9.1"  
21967..>48546  
/note="match: multiple ESTs; match: R21642 W46777 H45729  
W56058 R68369; match: R76745 H39623 H1045 W76544 H28751;  
match: W86968 N57451 H55328 W95130 R74257; match: H28751  
W46339 H02394 W95130 T49630; match: W32387 W76015; similar  
to FIBULIN-1 ISOFORM D PRECURSOR"  
join(<21967..22109,24433..24566,47763..47902)  
/gene="BR941F9.1"  
/codon\_start=2  
/product="BR941F9.1 (FIBULIN-1 ISOFORM D PRECURSOR LIKE)"  
/protein\_id="CAB08593.1"  
/db\_xref="pid:g1288141"  
/db\_xref="pid:g3087786"  
/db\_xref="gi:3087786"  
/translation="LQAKTIVRCIKSCRPNDVTCVDPVHTISLTPTREFT  
RDEIFLAIPTPPHSAQANITFDTEGLRDSFDIIRYMDGKMGVGVROVPIVG  
PFAVAVLKLENNYVGVGSHRNVNHIHSEYWF"  
22236..22536  
/note="AluX repeat: matches 301. .2 of consensus"  
22733..22735  
/gene="BR941F9.1"  
/note="clone CN38E2; CTA in this entry; substitution"  
/replace="cca"

```

variation      22866..22868
                /gene="BK941F9.1"
                /note="clone CN38E2; CAA in this entry; substitution"
                /replace="cga"
variation      23455..23457
                /gene="BK941F9.1"
                /note="clone CN38E2; CGC in this entry; substitution"
                /replace="cac"
variation      23592..23594
                /gene="BK941F9.1"
                /note="clone CN38E2; AAC in this entry; substitution"
                /replace="agc"
repeat_region  23873..23913
                /note="MIR repeat: matches 146..108 of consensus"
repeat_region  24154..24240
                /note="MIR repeat: matches 61..147 of consensus"
repeat_region  25946..26009
                /note="MIR repeat: matches 76..144 of consensus"
repeat_region  27964..28095
                /note="MIR repeat: matches 188..48 of consensus"
repeat_region  30253..30284
                /note="16 copies of 2 mer 84 & conserved"
repeat_region  31244..31279
                /note="6 copies of 6 mer 92 & conserved"
repeat_region  31943..32056
                /note="MIR repeat: matches 40..154 of consensus"
repeat_region  34691..34918
                /note="4 copies of 57 mer 99 & conserved"
repeat_region  36511..36574
                /note="2 copies of 32 mer 91 & conserved"
repeat_region  36606..36689
                /note="MIR repeat: matches 206..106 of consensus"
repeat_region  37266..37313
                /note="3 copies of 16 mer 88 & conserved"
variation      38069..38071
                /gene="BK941F9.1"
                /note="clone CN38E2; ATG in this entry; substitution"
                /replace="acg"
repeat_region  39230..39488
                /note="7 copies of 37 mer 99 & conserved"
repeat_region  39781..40083
                /note="AluX repeat: matches 302..1 of consensus"
variation      40235..40237
                /note="AluX repeat: matches 302..1 of consensus"

Query Match      83.2%; Score 15.8; DB 9; Length 127583;
Best Local Similarity 89.5%; Pred. No. 84;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 gatgattagcagagggt 19
        ||||| ||||| ||||
Db      42217 GATGATGACGACAGATGTT 42235

RESULT 14
HSDJ860P4      156791 bp      DNA      PRI      10-JUN-1999
LOCUS          Human DNA sequence from clone 860P4 on chromosome 20, complete
DEFINITION
ACCESSION      AL049594
NID            94757058
VERSION        AL049594.4 GI:4757058
KEYWORDS       HTG; CPG Island.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 156791)
AUTHORS       Matthews, L.
TITLE         Submitted (09-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL       CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
                requests: clonerequests@sanger.ac.uk
COMMENT       On May 6, 1999 this sequence version replaced gi:4680418.

```

## FEATURES

```

Source
1..156791
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="20"
  /clone_id="860P4"
  /clone_id="RPC15"
  /note="MIR repeat: matches 107..254 of consensus"
  /note="MIR repeat: matches 58..121 of consensus"
  /note="MIR repeat: matches 1..295 of consensus"
  /note="AluX repeat: matches 1833..1867"
  /note="7 copies 5 mer 99ag 86% conserved"
  /note="Charlieda repeat: matches 20..447 of consensus"
  /note="12 repeat: matches 2611..2730 of consensus"
  /note="15 copies 43 mer 57% conserved"
  /note="15 copies 43 mer 57% conserved"
  /note="308 copies 2 mer 99 56% conserved"
  /note="18 copies 18 mer 62% conserved"
  /note="14 copies 18 mer 69% conserved"
  /note="62 copies 2 mer tt 62% conserved"
  /note="MIR repeat: matches 16..532 of consensus"
  /note="MIR repeat: matches 702..1255 of consensus"
  /note="LMC repeat: matches 1293..1475 of consensus"
  /note="LMC repeat: matches 1..336 of consensus"
  /note="MER repeat: matches 1475..1773 of consensus"
  /note="AluX repeat: matches 158..306 of consensus"
  /note="LMC repeat: matches 1780..2569 of consensus"
  /note="LMC repeat: matches 1189..11481"
  /note="AluX repeat: matches 1..297 of consensus"
  /note="LMC repeat: matches 2569..3716 of consensus"
  /note="LMC repeat: matches 12593..13626"

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL.

This sequence is the entire insert of clone 860P4. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

860P4 is from the library RPC15 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcyrac2>.

```

misc_feature
/note="SVA repeat: matches 3. .955 of consensus"
13349. .14389
/note="Cpc Island"
evidence-not-experimental
13628. .14345
/note="SVA repeat: matches 684. .1386 of consensus"
14346. .14551
/note="L1MEC repeat: matches 3715. .3477 of consensus"
14553. .14705
/note="MER5B repeat: matches 25. .178 of consensus"
17313. .17726
/note="L1ME repeat: matches 5316. .5743 of consensus"
18943. .18976
/note="17 copies 2 mer ta 82% conserved"
19445. .19747
/note="ALUSq repeat: matches 1. .313 of consensus"
20812. .20959
/note="MER5A repeat: matches 6. .155 of consensus"
23451. .24088
/note="L1R9 repeat: matches 12. .625 of consensus"
25150. .25237
/note="MIR repeat: matches 27. .123 of consensus"
25549. .25866
/note="ALUtb repeat: matches 1. .305 of consensus"
27260. .27544
/note="ALUSx repeat: matches 1. .309 of consensus"
27545. .27606
/note="31 copies 2 mer ga 79% conserved"
27545. .27674
/note="5 copies 26 mer 68% conserved"
27566. .27737
/note="4 copies 43 mer 69% conserved"
27653. .27740
/note="22 copies 4 mer gga 91% conserved"
28338. .28417
/note="MIR repeat: matches 67. .150 of consensus"
28480. .28535
/note="14 copies 4 mer gatg 88% conserved"
28642. .28763
/note="MIR repeat: matches 137. .262 of consensus"
29025. .30142
/note="L1MC3 repeat: matches 6109. .7739 of consensus"
30176. .30213
/note="MADE1 repeat: matches 1. .38 of consensus"
30232. .30309
/note="39 copies 2 mer ta 71% conserved"
30238. .30309
/note="4 copies 18 mer 78% conserved"
30312. .30564
/note="L1MC3 repeat: matches 6354. .6609 of consensus"
30559. .30684
/note="7 copies 18 mer 71% conserved"
30694. .31030
/note="L1MC4 repeat: matches 6026. .6136 of consensus"
31560. .32061
/note="L1TIG repeat: matches 51. .512 of consensus"
32067. .32272
/note="MIR repeat: matches 12. .235 of consensus"
33507. .33834
/note="MSTB repeat: matches 1. .307 of consensus"
33835. .34128
/note="ALUSx repeat: matches 1. .295 of consensus"
34129. .34244
/note="MSTB repeat: matches 307. .426 of consensus"
35029. .35403
/note="MTR1A2 repeat: matches 1. .370 of consensus"
38035. .38304
/note="ALUSx repeat: matches 42. .309 of consensus"
38876. .39322
/note="L1TIG repeat: matches 13. .541 of consensus"
39958. .40252
/note="L1R16C repeat: matches 55. .384 of consensus"
40881. .41250

```

```

repeat_region
/note="MTR1B repeat: matches 1. .390 of consensus"
41579. .41798
/note="HAL1 repeat: matches 141. .376 of consensus"
41800. .41928
/note="3 copies 43 mer 76% conserved"
41841. .41908
/note="17 copies 4 mer gaag 91% conserved"
41927. .42194
/note="HAL1 repeat: matches 561. .829 of consensus"
42326. .42459
/note="HAL1 repeat: matches 1019. .1149 of consensus"
42532. .42908
/note="HAL1 repeat: matches 1336. .1712 of consensus"
42902. .43388
/note="L1ME3A repeat: matches 5594. .6117 of consensus"
44023. .44415
/note="MSTB repeat: matches 13. .426 of consensus"
46015. .46146
/note="MER3 repeat: matches 68. .205 of consensus"
46958. .47061
/note="L1 repeat: matches 3628. .3721 of consensus"
47331. .47625
/note="L1ME repeat: matches 5479. .5761 of consensus"
48491. .48848
/note="Tandem repeat. Tandem repeat region, contains one forced join. Digest suggests there may be 700bp missing from the repeat."
48501. .48842
/note="19 copies 18 mer 60% conserved"
48522. .48843
/note="161 copies 2 mer gg 58% conserved"
48549. .48838
/note="5 copies 58 mer 71% conserved"
48587. .48844
/note="6 copies 43 mer 63% conserved"
48647. .48854

```

```

Query Match      83.2% Score 15.8; DB 10; Length 156791;
Best Local Similarity 89.5%; Pred. No. 85;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 gatgattagcagagggt 19
||||| ||| |||||
Db 137058 GATGATTAGCAGAGGCTT 137076

```

```

RESULT 15
AC005895      87857 bp      DNA      PRI      31-OCT-1998
LOCUS      Homo sapiens chromosome 5, Bac clone 5m9 (LBNL H220), complete
DEFINITION
ACCESSION      AC005895
NID      G3818356
VERSION      AC005895.1 GI:3818356
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 87857)
AUTHORS      Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
Sequencing of human chromosome 5
Unpublished
JOURNAL      2 (bases 1 to 87857)
REFERENCE      Large Scale Sequence Analysis and Annotation with the Sequence
AUTHORS      Ricks,D.O.
TITLE      Comparison Analysis (SCAN) System
JOURNAL      Unpublished
REFERENCE      3 (bases 1 to 87857)
AUTHORS      Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,

```

TITLE  
JOURNAL  
COMMENT

Rojeski, H., Subramanian, S. and Martin, C. H.  
Direct Submission  
Submitted (31-Oct-1998) Human Genome Center, DOE Joint Genome  
Institute, Lawrence Berkeley National Laboratory, MS 74-157,  
Berkeley, CA 94720, U.S.A.  
Sequence submitted by:  
DOE Joint Genome Institute.

FEATURES  
Location/Qualifiers

Source

1. .87857  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="5p"  
/clone="Bac clone 5m9"  
/chromosome="5"  
/note="LBNL H220"  
complement(2700. .2820)  
/rpt\_family="MIR"  
2927. .3010  
/rpt\_family="MIR"  
8110. .8394  
complement(9814. .9871)  
/rpt\_family="MIR"  
11028. .11080  
/rpt\_family="MER44C"  
11464. .11674  
/rpt\_family="MER44C"  
complement(13364. .13408)  
/rpt\_family="MER8"  
13364. .13419  
/rpt\_family="Tigger2"  
13439. .13828  
/rpt\_family="Tigger2"  
complement(13924. .14028)  
/rpt\_family="MIR"  
14137. .14286  
complement(14711. .14884)  
/rpt\_family="MIR"  
complement(14953. .15095)  
/rpt\_family="Alu"  
16145. .16297  
/rpt\_family="MIR"  
17699. .17850  
/rpt\_family="MER45"  
19867. .19970  
/rpt\_family="MIR"  
19985. .20279  
complement(20345. .20597)  
/rpt\_family="Alu"  
complement(25131. .25215)  
/rpt\_family="MIR"  
complement(25381. .25669)  
/rpt\_family="Alu"  
26956. .27049  
/rpt\_family="MIR"  
27637. .27700  
/rpt\_family="MIR"  
27789. .28081  
/rpt\_family="Alu"  
31234. .31521  
/rpt\_family="Alu"  
31560. .31707  
/rpt\_family="MIR"  
complement(33494. .33661)  
/rpt\_family="MER45"  
34175. .34346  
/rpt\_family="MER42"  
complement(36898. .37135)  
/rpt\_family="MIR"  
complement(39739. .40026)  
/rpt\_family="Alu"

repeat\_region complement(40745. .40899)  
/rpt\_family="MIR"  
repeat\_region 42342. .43089  
/rpt\_family="MER3"  
repeat\_region complement(43177. .43466)  
/rpt\_family="Alu"  
repeat\_region complement(4865. .4973)  
/rpt\_family="Alu"  
repeat\_region 56084. .56361  
/rpt\_family="Alu"  
repeat\_region 57958. .58080  
/rpt\_family="MIR"  
repeat\_region complement(59837. .60108)  
/rpt\_family="Alu"  
repeat\_region complement(60303. .60380)  
/rpt\_family="MIR"  
repeat\_region 62346. .62473  
/rpt\_family="MIR"  
repeat\_region 62557. .62846  
/rpt\_family="Alu"  
repeat\_region complement(63914. .64067)  
/rpt\_family="MIR"  
repeat\_region 65798. .65999  
/rpt\_family="MIR"  
repeat\_region complement(67387. .67437)  
/rpt\_family="Alu"  
repeat\_region 68621. .68907  
/rpt\_family="Alu"  
repeat\_region complement(70187. .70357)  
/rpt\_family="MIR"  
repeat\_region 71198. .71489  
/rpt\_family="Alu"  
repeat\_region complement(75558. .75615)  
/rpt\_family="Alu"  
repeat\_region complement(75656. .75916)  
/rpt\_family="Alu"  
repeat\_region 76261. .76456  
/rpt\_family="MER20"  
repeat\_region 81116. .81408  
/rpt\_family="Alu"  
repeat\_region 82371. .82433  
/rpt\_family="MIR"  
repeat\_region 83142. .83302  
/rpt\_family="Alu"  
repeat\_region complement(83339. .83618)  
/rpt\_family="Alu"

BASE COUNT 20071 a 23584 c 24010 g 20192 t  
ORIGIN

Query Match 83.2%; Score 15.8; DB 11; Length 87857;  
Best Local Similarity 89.5%; Pred. No. 84;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 gatggattacgagggtt 19  
||||||| |||||  
DB 66996 GATGGATTGTCAGAGCGTT 67014

Search completed: August 18, 1999, 17:17:19  
Job time: 3485 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 17:18:31 ; Search time 148.09 seconds  
(without alignments)  
32.100 Million cell updates/sec

Title: US-09-004-395-4

Perfect score: 19  
Sequence: 1 gatgattacgagaggtt 19

Scoring table: IDENTITY\_NUC

Searched: 311585 segs, 125096042 residues

Database: N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	100.0	110000	1 X20248_01	Continuation (2 of
2	19	100.0	110000	1 X20248_02	Continuation (3 of
3	15.8	83.2	1320	1 Q80916	Plasmodium falciparum
4	15.8	83.2	110000	1 V21309_10	Continuation (11 of
5	15.4	81.1	3751	1 Q62461	Human embryonal kidney
6	15.4	81.1	4027	1 Q90972	Protein p140 CDNA
7	15.4	81.1	4027	1 Q90982	Protein p140 CDNA
8	15.4	81.1	3955	1 V22379	Human telomerase r
9	15.4	81.1	2176	1 V22380	Human telomerase r
10	15.4	81.1	3855	1 V22382	Human telomerase r
11	15.4	81.1	2171	1 V22426	Human telomerase r
12	15.4	81.1	4037	1 V22428	Human telomerase r
13	15.4	81.1	949	1 V27875	Human telomerase p
14	15.4	81.1	3787	1 V27876	Human telomerase p
15	15.4	81.1	4023	1 V60320	Human telomerase p
16	15.4	81.1	3346	1 V60321	Human telomerase p
17	15.4	81.1	2357	1 X15923	Alternatively spliced
18	15.4	81.1	2175	1 X15925	CDNA encoding a cca
19	15.4	81.1	4042	1 V72117	Nucleotide sequence
20	15.4	81.1	2089	1 V72126	Human catalytic te
21	14.8	77.9	709	1 Q03374	Partial sequence o
22	14.8	77.9	658	1 Q01809	Consensus sequence
23	14.8	77.9	11616	1 Q38105	Vector pHEBO-DR-LU
24	14.8	77.9	10580	1 Q38104	Vector pHEBO-DR-CA
25	14.8	77.9	2662	1 Q37757	PYMOS1. DNA sequen
26	14.8	77.9	17327	1 Q44278	Serglycin - proteo
27	14.8	77.9	658	1 Q44277	Serglycin - proteo
28	14.8	77.9	11616	1 Q67390	Construct pHEBO-DR
29	14.8	77.9	1556	1 Q95035	Rat hippocampal ne
30	14.8	77.9	1200	1 Q95036	Rat hippocampal ne
31	14.8	77.9	297	1 T26100	Human gene signatu
32	14.8	77.9	14985	1 T43574	Human box-dependen
33	14.8	77.9	14985	1 V15693	Human Bln1 gene. B
34	14.8	77.9	11770	1 V52305	Streptococcus pneu
35	14.8	77.9	110000	1 V21309_07	Continuation (8 of
36	14.8	77.9	23677	1 X16323	Partial human geno
37	14.4	75.8	378	1 Q61383	Human brain expres
38	14.4	75.8	110000	1 V21309_03	Continuation (4 of
39	14.2	74.7	5376	1 N90225	Malaria-specific p
40	14.2	74.7	5376	1 N90732	Sequence of plasmid
41	14.2	74.7	1254	1 N90733	Genomic DNA from p
42	14.2	74.7	1254	1 N92937	CDNA from Plasmid
43	14.2	74.7	950	1 Q28115	P.falciparum LSA-R

## ALIGNMENTS

44 14.2 74.7 1496 1 Q28119  
45 14.2 74.7 7902 1 V52245

P.falciparum LSA g  
Streptococcus pneu

RESULT 1  
X20248\_01/c  
Continuation (2 of 10) of X20248 from base 100001 (Borrelia burgdorferi polynucleotide  
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
WP Fragment Name Begin End  
WP X20248\_00 1 110000  
WP X20248\_01 100001 210000  
WP X20248\_02 200001 310000  
WP X20248\_03 300001 410000  
WP X20248\_04 400001 510000  
WP X20248\_05 500001 610000  
WP X20248\_06 600001 710000  
WP X20248\_07 700001 810000  
WP X20248\_08 800001 910000  
WP X20248\_09 900001 910715

Query Match 100.0%; Score 19; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.97;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gatgattacgagaggtt 19  
DB 105074 GATGATTACGAGAGGTT 105056

RESULT 2  
X20248\_02/c  
Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide  
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
WP Fragment Name Begin End  
WP X20248\_00 1 110000  
WP X20248\_01 100001 210000  
WP X20248\_02 200001 310000  
WP X20248\_03 300001 410000  
WP X20248\_04 400001 510000  
WP X20248\_05 500001 610000  
WP X20248\_06 600001 710000  
WP X20248\_07 700001 810000  
WP X20248\_08 800001 910000  
WP X20248\_09 900001 910715

Query Match 100.0%; Score 19; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.97;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gatgattacgagaggtt 19  
DB 5074 GATGATTACGAGAGGTT 5056

RESULT 3  
Q80916  
ID Q80916 standard; cDNA; 1320 BP.  
AC Q80916:  
DE 28-AUG-1995 (first entry)  
KW Plasmodium falciparum LSA-1 repeatless gene cDNA.  
KW Plasmodium falciparum LSA-1 repeatless gene; recombinant poxvirus;  
KW multicomponent multistage malarial vaccines; immunogens;  
OS malaria diagnosis; ss.  
PN W09428930-A.  
PD 22-DEC-1994.  
PF 10-JUN-1994; U06652.  
PR 11-JUN-1993; US-075783.

PR 09-JUN-1994: US-25707.  
 PA (VIR0-) VIR0GENETICS CORP..  
 PI De Taisne C, Paoletti E, Tine JA;  
 DR WPJ: 95-036113/05.  
 PT Recombinant poxvirus contg. Plasmodium DNA in non-essential  
 PT region - useful in vaccines against malaria and for prodn. of  
 PT Plasmodium immunogens  
 PS Claim 3: Fig 11, 183pp: English.  
 CC 080916 is the P. falciparum LSA-1 repeatless gene cDNA sequence.  
 CC New recombinant poxviruses containing either the SRA, ABRA, Phsp7  
 CC ANA-1, PFS25, PFS16, CSP, PFS32, LSA-1, LSA-1 repeatless, MSA-1,  
 CC MSA-1 (N-terminal p83 or C-terminal 942) genes, or a combination  
 CC of these in non-essential regions of their genomes are claimed.  
 CC These poxviruses (pref. with a virulence reducing genomic  
 CC deletion or disruption) can be used as vaccines against malaria  
 CC and for the prodn. of Plasmodium immunogens. These viruses  
 CC provide multicomponent, multistage vaccines due to their expression  
 CC of sporozoite, liver stage, blood stage and sexual stage proteins.  
 CC Sequence 1320 Bp; 619 A; 128 C; 227 G; 346 T;

Query Match	83.2%	Score 15.8	DB 1	Length 1320
Best Local Similarity	89.5%	Pred No. 25		
Matches 17; Conservative	0	Mismatches 2	Indels 0	Gaps 0

```
QY      1 gatgattagcagggtt 19
          ||| ||| ||| ||| |||
Db      526 GATGTATTAGCAGAGGATT 544
```

RESULT 4  
V21209\_10  
Continuation (11 of 17) of V21209 from base 1000001 (Methanococcus jannaschii circular c  
MP Sequence split into 17 fragments LOCUS V21209 Accession V21209

Query Match	83.2%	Score 15.8	DB 1	Length 110000
Best Local Similarity	89.5%	Pred. No. 39		
Matches 17	Conservative	0	Mismatches 2	Indels 0
				Gaps 0

```

QY      1 gatgattagcagaggtt 19
          |||||  |||  |||||
Db 15043 GATGATAAGCATAGGTT 15061

```

RESULT 5  
ID 062461  
AC 062461 standard; cDNA; 3751 BP.  
DT 09-NOV-1994 (first entry)  
DE Human embryonal kinase 2 receptor.  
KW Embryonal kinase; HEK; protein tyrosine kinase; PTK; tumour;  
KW cancer; therapy; amplification; primer; polymerase chain reaction;  
KW PCR; ss.  
OS Homo sapiens.

EH	Key	Location/Qualifiers
FT	cds	1..2973
FT		/tag= a
PN		DE4233782-A.
PD		14-APR-1994.
PF		07-OCT-1992; 233782.
PR		07-OCT-1992; DE-233782.
PA		(CHEM-) CHEMOTHERAPEUTISCHES FORSCHUNG.
PI		Holtrich U, Ruebsamen-Waigmann H, Streibhardt K;
DR		WEI; 94-127194/16.
DR		P-PSDB; R51899.
PT		Human embryonal kinase 2-receptor protein - useful in tumour
PT		diagnosis and therapy
PS		Claim 4: Page 7-10; 11pp; German.
CC		RNA from human embryonic tissue was isolated. With the use of
CC		primer P6(4) pRK-specific cDNA was synthesised. The cDNA was
CC		amplified using primers P6(4) and NS. A 2097 bp DNA fragment was
CC		obtained. Primers E5, F12 and E6 were then used in the isolation of
CC		the C-terminal of the HK2 receptor gene.
Q0		Sequence 3751 bp; 739 A; 1150 C; 1123 G; 739 T;

Query Match	81.1%	Score 15.4	DB 1	Length 3751
Best Local Similarity	94.1%	Pred. No. 45		
Matches 16	Conservative 0	Mismatches 1	Indels 0	Gaps 0

```

OY      1 gatgattacgaggg 17
          |||||
Db      3168 GATGATTAGGAGAGG 3184

```

RESULT 6  
 Q09072  
 ID Q09072 standard; cDNA to mRNA; 4027 BP.  
 AC Q09072;  
 DT 24-NOV-1995 (first entry)  
 DE Protein p140 cDNA from rat skeletal muscle myoblast cell line L6.  
 KW Protein p140; Insulin; tyrosine phosphorylation; ss.  
 OS Rattus rattus.  
 PN EP-659883-A.  
 PD 28-JUN-1995.  
 PE 24-NOV-1994; 118524.  
 PR 24-NOV-1993; JP-315806.  
 PI (ONOGU) ONO PHARM CO LTD.  
 PI Kitagawa K, Ohno H, Tajima H;  
 DR WPI; 95-226291/30.  
 FT Isolated protein p140 polypeptide - and treatment of diabetes based  
 PT on tyrosine phosphorylation of protein p140.  
 PS Claim 5; Page 25-27; 42pp; English.  
 CC A cDNA library was established from rat skeletal myoblast cell line  
 CC L6. DNA fragments of approx. 400 bp were recovered and subjected to  
 CC cloning. 20 plasmids were sequenced. Sequence data of cDNA fragments  
 CC were constructed to the linkage sequences with the DNA programme  
 CC DNASIS. The basic sequence portrayed in Q09072 was hence constructed.  
 CC From sequence data of the whole cDNA length, the ORF was determined.  
 CC The AA sequence was further translated and the sequence thus  
 CC established is illustrated in R75843. One of the firms possesses  
 CC the 2993-bp ORF, that was approximated to 3000 bp of the whole ORF  
 CC length of the Eck family.  
 Q0 Sequence 4027 BP; 885 A; 1141 C; 1166 G; 835 T;

Query Match	81.1%	Score 15.4;	DB 1;	Length 4027;
Best Local Similarity	94.1%;	Pred. NO. 45;		
Matches 16;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

```
QY      1 gatgattagcaggg 17
          |||||
Db 3433 GATGATTAGAGAGGG 344
```

RESULT  
Q90982



ID Q09082 standard; cDNA to mRNA; 4027 BP.  
AC Q09082;  
DT 24-NOV-1995 (first entry)  
DE Protein p140 cDNA from rat skeletal muscle myoblast cell line L6.  
KW Protein p140; Insulin; tyrosine phosphorylation; ss.  
OS Rattus rattus.  
FH Key Location/Qualifiers  
FT cds 262..3243  
FT /\*tag= a  
PN EP-659883-A.  
PD 28-JUN-1995.  
PR 24-NOV-1994; 118524.  
PR 24-NOV-1993; JP-315806.  
PA (ONOY) ONO PHARM CO LTD.  
PI Kitagawa K, Ohno H, Tajima H;  
DR P-PSDB: R75844.  
PT Isolated protein p140 polypeptide - and treatment of diabetes based  
PT on tyrosine phosphorylation of protein p140.  
PS Example; Page 28-33; 42pp; English.  
CC Q09082 does not appear to be referred to in the spec. although  
CC it is listed in the sequence listings (SEQ ID No. 4). It is the  
CC same length as SEQ ID No. 3 and may be identical to it (see  
CC Q090972).  
SO Sequence 4027 BP; 885 A; 1141 C; 1166 G; 835 T;  
  
Query Match 81.1%; Score 15.4; DB 1; Length 4027;  
Best Local Similarity 94.1%; Pred. No. 45;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 gatgattagcagaagg 17  
DB 3433 GATGATTAGAGAGG 3449  
|||||  
RESULT 8  
ID V22379/c  
AC V22379; standard; cDNA; 3955 BP.  
DT 13-AUG-1998 (first entry)  
DE Human telomerase reverse transcriptase encoding cDNA.  
KW Human; telomerase reverse transcriptase; hTERT; TR; diagnosis;  
KW prognosis; cell proliferation; cancer; ageing; ribonucleoprotein; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 56..3454  
FT /\*tag= a  
FT /product= "telomerase reverse transcriptase"  
FT  
FT  
FT  
PN GB3317891-A.  
PD 08-APR-1998.  
PE 01-OCT-1997; 020890.  
PR 14-AUG-1997; US-915503.  
PR 01-OCT-1996; US-724643.  
PR 18-APR-1997; US-844419.  
PR 25-APR-1997; US-846017.  
PR 06-MAY-1997; US-851843.  
PR 09-MAY-1997; US-854050.  
PR 14-AUG-1997; US-911312.  
PR 14-AUG-1997; US-912951.  
PA (GERO-) GERON CORP.  
PA (UYTE-) UNIV TECHNOLOGY CORP.  
PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,  
PI Morin GB, Nakamura T, Harley CB;  
DR P-PSDB: W46937.  
DR P-PSDB: W46937.  
PT Pure and recombinant human Telomerase Reverse Transcriptase and its  
PT variants - are useful in the diagnosis, prognosis and treatment of  
PT cell proliferation conditions especially cancer and ageing  
PS Claim 5; Fig 16; 387pp; English.  
CC The present sequence encodes human telomerase reverse transcriptase  
CC (hTERT), which is a ribonucleoprotein. The present invention also  
CC describes the following methods: (A) determining whether a test compound

CC is a modulator of hTERT, by detecting the change in hTERT recombinant  
CC protein or polynucleotide, on administration of the compound; (B)  
CC preparation of recombinant telomerase by contacting a protein  
CC preparation of hTERT with a telomerase RNA component; (C) detection of  
CC the hTERT RNA or protein in a sample by binding a relevant probe to the  
CC sample and detecting the complex formed or in the case of RNA detection,  
CC amplifying the product and correlating the presence of complex or  
CC increasing the product with presence of hTERT in the sample; and (D)  
CC increasing the proliferation of a vertebrate cell by increasing hTERT  
CC expression; and (E) the use of an agent that causes an increase in cell  
CC vertebrate cell proliferation to create a medicament that inhibits  
CC ageing. A protein preparation of hTERT and the polynucleotide encoding  
CC hTERT can be used in the manufacture of medicaments for inhibiting the  
CC effect of ageing or cancer. Inhibitors of telomerase activity can be  
CC used to treat conditions that are associated with high telomerase  
CC activity. A protein preparation of hTERT can also be used in the new  
CC methods.  
SO Sequence 3955 BP; 648 A; 1353 C; 1251 G; 703 T;  
  
Query Match 81.1%; Score 15.4; DB 1; Length 3955;  
Best Local Similarity 94.1%; Pred. No. 45;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 gatgattagcagaagg 17  
DB 3193 GATGATTAGCAGAGG 3177  
|||||  
RESULT 9  
ID V22380/c  
AC V22380; standard; cDNA; 2176 BP.  
DT 13-AUG-1998 (first entry)  
DE Human telomerase reverse transcriptase cDNA clone 712562.  
KW Human; telomerase reverse transcriptase; hTERT; TR; diagnosis;  
KW prognosis; cell proliferation; cancer; ageing; ribonucleoprotein; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 23..802  
FT /\*tag= a  
FT /product= "telomerase reverse transcriptase"  
FT  
FT  
FT  
PN GB3317891-A.  
PD 08-APR-1998.  
PE 01-OCT-1997; 020890.  
PR 14-AUG-1997; US-915503.  
PR 01-OCT-1996; US-724643.  
PR 18-APR-1997; US-844419.  
PR 25-APR-1997; US-846017.  
PR 06-MAY-1997; US-851843.  
PR 09-MAY-1997; US-854050.  
PR 14-AUG-1997; US-911312.  
PR 14-AUG-1997; US-912951.  
PA (GERO-) GERON CORP.  
PA (UYTE-) UNIV TECHNOLOGY CORP.  
PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,  
PI Morin GB, Nakamura T, Harley CB;  
DR P-PSDB: W46938.  
DR P-PSDB: W46938.  
PT Pure and recombinant human Telomerase Reverse Transcriptase and its  
PT variants - are useful in the diagnosis, prognosis and treatment of  
PT cell proliferation conditions especially cancer and ageing  
PS Example 1; Fig 18; 387pp; English.  
CC The present sequence represents a human telomerase reverse transcriptase  
CC (hTERT) cDNA clone from the present invention. The present invention also  
CC describes the following methods: (A) determining whether a test compound  
CC is a modulator of hTERT, by detecting the change in hTERT recombinant  
CC protein or polynucleotide, on administration of the compound; (B)  
CC preparation of recombinant telomerase by contacting a protein  
CC preparation of hTERT with a telomerase RNA component; (C) detection of  
CC the hTERT RNA or protein in a sample by binding a relevant probe to the  
CC sample and detecting the complex formed or in the case of RNA detection,  
CC amplifying the product and correlating the presence of complex or

CC amplification product with presence of hTERT in the sample; and (D)  
 CC increasing the proliferation of a vertebrate cell by increasing hTERT  
 CC expression; and (E) the use of an agent that causes an increase in cell  
 CC vertebrate cell proliferation to create a medicament that inhibits  
 CC ageing. A protein preparation of hTERT and the polynucleotide encoding  
 CC hTERT can be used in the manufacture of medicaments for inhibiting the  
 CC effect of ageing or cancer. Inhibitors of telomerase activity can be  
 CC used to treat conditions that are associated with high telomerase  
 CC activity. A protein preparation of hTERT can also be used in the new  
 CC methods.

SQ Sequence 2176 BP; 432 A; 679 C; 641 G; 422 T;

Query Match 81.1%; Score 15.4; DB 1; Length 2176;  
 Best Local Similarity 94.1%; Pred. No. 42;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatgattagcagagg 17  
 ||||| |||||  
 DB 1334 GATGAGTAGCAGAGGG 1318

RESULT 10  
 V22382/c  
 ID V22382 standard; cDNA; 3855 BP.

AC V22382;  
 DT 13-AUG-1998 (first entry)  
 DE Human telomerase reverse transcriptase Delta182 variant encoding cDNA.  
 KW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;  
 KW prognosis; cell proliferation; cancer; ageing; ribonucleoprotein; ss.  
 OS Synthetic.  
 OS Homo sapiens.

FT Key Location/Qualifiers  
 FT CDS 56..2479  
 FT /tag= a  
 FT /product= "telomerase reverse transcriptase variant"

PN GB2317891-A.  
 PD 08-APR-1998.  
 PF 01-OCT-1997; 020890.  
 PR 14-AUG-1997; US-915503.  
 PR 01-OCT-1996; US-724643.  
 PR 18-APR-1997; US-844419.  
 PR 25-APR-1997; US-846017.  
 PR 06-MAY-1997; US-851843.  
 PR 09-MAY-1997; US-854050.  
 PR 14-AUG-1997; US-911312.  
 PR 14-AUG-1997; US-912951.

PA (GERO-) GERON CORP.  
 PI (UYTE-) UNIV TECHNOLOGY CORP.  
 PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,  
 PI Morin GB, Nakamura T, Harley CB;  
 DR WPI: 98-171633/16.  
 DR P-PSDB: W46997.

PT Pure and recombinant human Telomerase Reverse Transcriptase and its  
 PT variants - are useful in the diagnosis, prognosis and treatment of  
 PT cell proliferation conditions especially cancer and ageing  
 PS Disclosure: Fig 20; 387pp; English.  
 CC The present sequence encodes a human telomerase reverse transcriptase  
 CC (hTERT) variant from the present invention. The present invention also  
 CC describes the following methods: (A) determining whether a test compound  
 CC is a modulator of hTERT, by detecting the change in hTERT recombinant  
 CC protein or polynucleotide, on administration of the compound; (B)  
 CC preparation of recombinant telomerase by contacting a protein  
 CC preparation of hTERT with a telomerase RNA component; (C) detection of  
 CC the hTERT RNA or protein in a sample by binding a relevant probe to the  
 CC sample and detecting the complex formed or in the case of RNA detection,  
 CC amplifying the product and correlating the presence of complex or  
 CC amplification product with presence of hTERT in the sample; and (D)  
 CC increasing the proliferation of a vertebrate cell by increasing hTERT  
 CC expression; and (E) the use of an agent that causes an increase in cell  
 CC vertebrate cell proliferation to create a medicament that inhibits  
 CC ageing. A protein preparation of hTERT and the polynucleotide encoding  
 CC hTERT can be used in the manufacture of medicaments for inhibiting the

CC effect of ageing or cancer. Inhibitors of telomerase activity can be  
 CC used to treat conditions that are associated with high telomerase  
 CC activity. A protein preparation of hTERT can also be used in the new  
 CC methods.

SQ Sequence 3855 BP; 651 A; 1300 C; 1226 G; 678 T;

Query Match 81.1%; Score 15.4; DB 1; Length 3855;  
 Best Local Similarity 94.1%; Pred. No. 45;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatgattagcagagg 17  
 ||||| |||||  
 DB 3011 GATGAGTAGCAGAGGG 2995

RESULT 11  
 V22426/c  
 ID V22426 standard; DNA; 2171 BP.

AC V22426;  
 DT 13-AUG-1998 (first entry)  
 DE Human telomerase reverse transcriptase 63 kDa clone 712562.  
 KW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;  
 KW prognosis; cell proliferation; cancer; ageing; ribonucleoprotein; ss.  
 OS Synthetic.  
 OS Homo sapiens.

FT Key Location/Qualifiers  
 FT CDS 223..1713  
 FT /tag= a  
 FT /product= "telomerase reverse transcriptase 63 kDa"  
 FT /transl\_except= (pos:325..327,aa:Xaa)  
 FT /note= "Xaa= unspecified"

PN GB2317891-A.  
 PD 08-APR-1998.  
 PF 01-OCT-1997; 020890.  
 PR 14-AUG-1997; US-915503.  
 PR 01-OCT-1996; US-724643.  
 PR 18-APR-1997; US-844419.  
 PR 25-APR-1997; US-846017.  
 PR 06-MAY-1997; US-851843.  
 PR 09-MAY-1997; US-854050.  
 PR 14-AUG-1997; US-911312.  
 PR 14-AUG-1997; US-912951.

PA (GERO-) GERON CORP.  
 PI (UYTE-) UNIV TECHNOLOGY CORP.  
 PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,  
 PI Morin GB, Nakamura T, Harley CB;  
 DR WPI: 98-171633/16.  
 DR P-PSDB: W56109.

PT Pure and recombinant human Telomerase Reverse Transcriptase and its  
 PT variants - are useful in the diagnosis, prognosis and treatment of  
 PT cell proliferation conditions especially cancer and ageing  
 PS Example 1; Fig 68; 387pp; English.  
 CC The present sequence represents a human telomerase reverse transcriptase  
 CC (hTERT) DNA clone from the present invention. The present invention also  
 CC describes the following methods: (A) determining whether a test compound  
 CC is a modulator of hTERT, by detecting the change in hTERT recombinant  
 CC protein or polynucleotide, on administration of the compound; (B)  
 CC preparation of recombinant telomerase by contacting a protein  
 CC preparation of hTERT with a telomerase RNA component; (C) detection of  
 CC the hTERT RNA or protein in a sample by binding a relevant probe to the  
 CC sample and detecting the complex formed or in the case of RNA detection,  
 CC amplifying the product and correlating the presence of complex or  
 CC amplification product with presence of hTERT in the sample; and (D)  
 CC increasing the proliferation of a vertebrate cell by increasing hTERT  
 CC expression; and (E) the use of an agent that causes an increase in cell  
 CC vertebrate cell proliferation to create a medicament that inhibits  
 CC ageing. A protein preparation of hTERT and the polynucleotide encoding  
 CC hTERT can be used in the manufacture of medicaments for inhibiting the  
 CC effect of ageing or cancer. Inhibitors of telomerase activity can be  
 CC used to treat conditions that are associated with high telomerase  
 CC activity. A protein preparation of hTERT can also be used in the new  
 CC methods.

SQ Sequence 2171 BP; 433 A; 668 C; 640 G; 429 T;

Query Match  
Best Local Similarity 81.1%; Score 15.4; DB 1; Length 2171;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gatgattacagagg 17  
||||| |||||||  
Db 1330 GATGAGTACAGAGG 1314

RESULT 12  
V22428/c  
ID V22428 standard; CDNA; 4037 BP.  
AC V22428;  
DT 13-AUG-1998 (first entry)  
DE Human telomerase reverse transcriptase encoding CDNA refined sequence.  
KW Human; telomerase reverse transcriptase; hTERT; RT; diagnosis;  
OS Homo sapiens; cell proliferation; cancer; aging; ribonucleoprotein; ss.  
FH Key location/Qualifiers  
FT CDS 56..3454  
FT /tag- a "telomerase reverse transcriptase"  
FT /product- "refined sequence"

GB317891-A.  
PD 08-APR-1998.  
PF 01-OCT-1997; 020890.  
PR 14-AUG-1997; US-915503.  
PR 01-OCT-1996; US-724643.  
PR 18-APR-1997; US-844419.  
PR 25-APR-1997; US-846017.  
PR 05-MAY-1997; US-851843.  
PR 09-MAY-1997; US-854050.  
PR 14-AUG-1997; US-911312.  
PR 14-AUG-1997; US-912951.  
PA (GERO-) GERON CORP.  
PA (UYTE-) UNIV TECHNOLOGY CORP.  
PI Andrews WH, Czech TR, Chapman KB, Harley C, Langner J,  
PI Morin GB, Nakamura T, Harley CB;  
DR WPI; 98-171633/16.  
DR P-PSD; W56113.  
PT Pure and recombinant human Telomerase Reverse Transcriptase and its  
PT variants - are useful in the diagnosis, prognosis and treatment of  
PT cell proliferation conditions especially cancer and aging  
PS Example 1: Fig 74; 387pp; English.  
CC The present sequence encodes human telomerase reverse transcriptase  
CC (hTERT), which is a ribonucleoprotein. The present invention also  
CC describes the following methods: (A) determining whether a test compound  
CC is a modulator of hTERT, by detecting the change in hTERT recombinant  
CC protein or polynucleotide, on administration of the compound; (B)  
CC preparation of recombinant telomerase by contacting a protein  
CC preparation of hTERT with a telomerase RNA component; (C) detection of  
CC the hTERT RNA or protein in a sample by binding a relevant probe to the  
CC sample and detecting the complex formed or in the case of RNA detection,  
CC amplifying the product and correlating the presence of complex or  
CC amplification product with presence of hTERT in the sample; and (D)  
CC increasing the proliferation of a vertebrate cell by increasing hTERT  
CC expression; and (E) the use of an agent that causes an increase in cell  
CC vertebrate cell proliferation to create a medicament that inhibits  
CC aging. A protein preparation of hTERT and the polynucleotide encoding  
CC hTERT can be used in the manufacture of medicaments for inhibiting the  
CC effect of aging or cancer. Inhibitors of telomerase activity can be  
CC used to treat conditions that are associated with high telomerase  
CC activity. A protein preparation of hTERT can also be used in the new  
CC methods.  
SQ Sequence 4037 BP; 682 A; 1362 C; 1275 G; 714 T;

Query Match  
Best Local Similarity 81.1%; Score 15.4; DB 1; Length 4037;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gatgattacagagg 17  
||||| |||||||  
Db 3193 GATGAGTACAGAGG 3177

RESULT 13  
V27875/c  
ID V27875 standard; CDNA; 949 BP.  
AC V27875;  
DT 12-OCT-1998 (first entry)  
DE Human telomerase protein 2 (TP2) 3' CDNA partial clone TP2-15.  
KW TP2; human; telomerase protein 2; cancer; AIDS; aging; therapy;  
OS ss.  
OS Homo sapiens.  
PN MO9821343-A1.  
PD 22-MAY-1998.  
PF 13-NOV-1997; U21248.  
PR 16-OCT-1997; US-951733.  
PR 15-NOV-1996; US-871189.  
PR 11-JUN-1997; US-873039.  
PA (AMGE-) AMGEN CANADA INC.  
PA (AMGE-) AMGEN INC.  
PI Harrington LA, Robinson MO;  
PI WPI; 98-297946/26.  
DR New nucleic acid encoding human telomerase protein-2 - used for  
PT regulating telomerase activity, e.g. for treating cancer or acquired  
PT immune deficiency syndrome  
PS Example 5; Fig 7; 150pp; English.  
CC Partial CDNA clone TP2-15 comprises the 3' region of human  
CC telomerase protein 2 (TP2) CDNA, 133 bases of which overlap with  
CC the 3' end of another partial clone, #32 (see V27872). It was  
CC isolated from a human colon tumor cell line LIM1863 CDNA library  
CC using an 830 bp XhoI fragment of clone #32 as probe. Clones TP2-15  
CC and #32 have been combined to provide a full-length sequence (see  
CC V27876) coding for human TP2 (see W61350). Expressing TP2 genes  
CC in a cell is used to increase telomerase activity and thus  
CC proliferation for treatment of e.g. HIV infection, AIDS and  
CC aging disorders, while expressing an inactive mutant of TP2 (or  
CC molecule antisense to the gene) is used to decrease telomerase  
CC activity, e.g. for treatment of cancer. TP2 polypeptides can also  
CC be used to screen for agents that inhibit TP2 activity or its  
CC binding to TRIP1 (see W61347) or telomerase RNA, potentially  
CC useful therapeutically, also to raise specific antibodies useful  
CC in immunoassays and therapeutically as inhibitors. Nucleic acid  
CC fragments are used as diagnostic probes for detecting/quantifying  
CC TP2 DNA. Also contemplated are transgenic animals in which the TP2  
CC gene has been inactivated or is overexpressed. TP2 polypeptides  
CC are administered i.v., s.c. or orally, or they are delivered from  
CC engineered cells or gene therapy vectors.  
SQ Sequence 949 BP; 176 A; 331 C; 268 G; 174 T;

Query Match  
Best Local Similarity 81.1%; Score 15.4; DB 1; Length 949;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gatgattacagagg 17  
||||| |||||||  
Db 356 GATGAGTACAGAGG 340

RESULT 14  
V27876/c  
ID V27876 standard; CDNA; 3798 BP.  
AC V27876;  
DT 12-OCT-1998 (first entry)  
DE Human telomerase protein 2 (TP2) full-length CDNA.  
KW TP2; human; telomerase protein 2; cancer; AIDS; aging; therapy;  
OS ss.  
OS Homo sapiens.  
FH Key location/Qualifiers  
FT CDS 2..3466

```

FT WO9821343-A1. /*tag= a
PN 22-MAY-1998.
PF 13-NOV-1997; U21248.
PR 16-OCT-1997; US-951733.
PR 15-NOV-1996; US-871189.
PR 11-JUN-1997; US-873039.
PA (AMGE-) AMGEN CANADA INC.
PA (AMGE-) AMGEN INC.
PI Harrington LA, Robinson MO;
DR WPI: 98-297946/26.
DR P-PSDB: W27876.
PT New nucleic acid encoding human telomerase protein-2 - used for
PT regulating telomerase activity, e.g. for treating cancer or acquired
PT immune deficiency syndrome
PS Claim 1c; Fig 8; 150pp; English.
CC This full-length cDNA clone codes for a human telomerase protein
CC 2 (TP2, see W61350), a novel protein of the telomerase complex.
CC The sequence was deduced from overlapping partial clones #32 (see
CC V27872) and TP2-15 (see V27875), which were obtained from a human
CC colon tumour cell line LIM1863 cDNA. Expressing TP2 genes in a
CC cell is used to increase telomerase activity and thus proliferation
CC for treatment of e.g. HIV infection, AIDS and ageing disorders,
CC while expressing an inactive mutant of TP2 (or molecule antisense
CC to the gene) is used to decrease telomerase activity, e.g. for
CC treatment of cancer. TP2 polypeptides can also be used to screen
CC for agents that inhibit TP2 activity or its binding to TRIP1 (see
CC W61347) or telomerase RNA, potentially useful therapeutically,
CC also to raise specific antibodies useful in immunoassays and
CC therapeutically as inhibitors. Nucleic acid fragments are used
CC as diagnostic probes for detecting/quantifying TP2 DNA. Also
CC contemplated are transgenic animals in which the TP2 gene has been
CC inactivated or is overexpressed. TP2 polypeptides are administered
CC i.v., s.c. or orally, or they are delivered from engineered cells
CC or gene therapy vectors.
SQ Sequence 3798 BP; 613 A; 1311 C; 1212 G; 662 T;

Query Match 81.1%; Score 15.4; DB 1; Length 3798;
Best Local Similarity 94.1%; Pred. No. 45;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatgattagcagagg 17
   ||||| |||||
DB 3205 GATGAGTAGCAGAGCG 3189

RESULT 15
V60320/c
ID V60320 standard; cDNA: 4023 BP.
AC V60320;
DR 04-DEC-1998 (first entry)
DE Human telomerase gene referred to as hEST2.
KW Catalytic subunit; human; telomerase; telomere maintenance;
KW diagnosis; treatment; cancer; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 59..3458
FT /*tag= a
PN WO9837181-A2.
PD 27-AUG-1998.
PF 20-FEB-1998; U03404.
PR 30-OCT-1997; US-064322.
PR 20-FEB-1997; US-038750.
PR 20-MAY-1997; US-047151.
PR 01-AUG-1997; US-054549.
PR 14-AUG-1997; US-055762.
PA (WHD ) WHITEHEAD INST BIOMEDICAL RES.
PI Counter CM, Meyerson M, Weinberg RA;
DR WPI: 98-495367/42.
DR P-PSDB: W71376.
PT New isolated human telomerase catalytic sub-unit gene - used to
PT develop products for increasing or reducing the life span of cells

```

```

PT such as cancer cells or transformed cells
PS Claim 5; Fig 5A-B; 96pp; English.
CC The present sequence encodes the catalytic subunit of a human
CC telomerase holoenzyme. Disruption of the telomerase gene alters
CC telomere maintenance. The DNA is essential for telomerase activity,
CC and the protein is physically associated with telomerase and a
CC constituent of active telomerase complex. The products can be used
CC for increasing or reducing the lifespan of cells such as cancer cells
CC or transformed cells. They can also be used in the diagnosis and
CC treatment of malignancies. In addition, cells with a longer lifespan
CC can be transplanted into or grafted onto an individual (e.g. as skin
CC grafts), as systems for delivery of therapeutic proteins, such as hormones
CC and enzymes), to whom they provide therapeutic benefit.
SQ Sequence 4023 BP; 668 A; 1363 C; 1277 G; 715 T;

Query Match 81.1%; Score 15.4; DB 1; Length 4023;
Best Local Similarity 94.1%; Pred. No. 45;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatgattagcagagg 17
   ||||| |||||
DB 3196 GATGAGTAGCAGAGCG 3180

```

Search completed: August 18, 1999, 17:18:36  
Job time: 3268 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 17:04:42 ; Search time 767.16 Seconds  
(without alignments)  
48.853 Million cell updates/sec

Title: US-09-004-395-4

Perfect score: 19  
Sequence: 1 gatgagtagacagagggt 19

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*  
44: gb\_est25: \*  
45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: em\_est20: \*  
53: em\_est21: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	16.4	86.3	479	21	T64274	T64274 yc09b09.r1
C 2	16.4	86.3	499	22	R59018	R59018 y996c08.r1
C 3	16.4	86.3	374	24	H61749	H61749 yu39a09.r1
C 4	16.4	86.3	239	33	AA393708	AA393708 zc71g10.r
5	15.8	83.2	309	22	H01442	H01442 y199b09.s1
6	15.8	83.2	420	45	A1353711	A1353711 zeb0888.s
7	15.4	81.1	407	20	T09332	T09332 EST07225.1n
8	15.4	81.1	408	21	T79293	T79293 yd70g04.r1
9	15.4	81.1	461	22	R50053	R50053 yj59b10.r1
10	15.4	81.1	267	22	R58460	R58460 G3742 fetal
C 11	15.4	81.1	446	27	AA002971	AA002971 mg50e04.r
C 12	15.4	81.1	278	28	AA090488	AA090488 y0582.seq
C 13	15.4	81.1	560	30	AA241858	AA241858 mk21c11.r
C 14	15.4	81.1	342	30	AA244838	AA244838 mk13b09.r
C 15	15.4	81.1	188	30	AA261440	AA261440 m87f02.r
C 16	15.4	81.1	303	31	AA299878	AA299878 EST12462
C 17	15.4	81.1	428	32	AA350990	AA350990 EST58583
18	15.4	81.1	379	34	AA455591	AA455591 aa17c09.r
19	15.4	81.1	485	37	AA683854	AA683854 vt05f08.r
C 20	15.4	81.1	459	41	A109440	A109440 ud37b02.r
C 21	15.4	81.1	683	42	A1153237	A1153237 uc51b04.r
C 22	15.4	81.1	403	42	A1153378	A1153378 uc52c09.r
C 23	15	78.9	280	20	D26712	D26712 CELK01D3R
24	15	78.9	302	31	AA327558	AA327558 EST30914
25	15	78.9	628	47	A1485028	A1485028 EST243308
C 26	15	78.9	900	50	A0067367	A0067367 A0067367
27	14.8	77.9	327	24	D65439	D65439 CELK056E9R
28	14.8	77.9	495	24	N33046	N33046 y05e01.s1
29	14.8	77.9	455	25	N99752	N99752 SWACAS69SK
30	14.8	77.9	313	26	W25334	W25334 zb89f02.r1
31	14.8	77.9	127	26	W25453	W25453 zb90g05.r1
32	14.8	77.9	393	27	AA046218	AA046218 zkr7e08.r
33	14.8	77.9	380	27	C01356	C01356 HUMGS00833
34	14.8	77.9	429	29	AA149480	AA149480 z127f02.r
35	14.8	77.9	312	31	AA304388	AA304388 EST17140
36	14.8	77.9	272	31	AA310196	AA310196 EST18103
37	14.8	77.9	400	31	AA328948	AA328948 EST32774
38	14.8	77.9	355	32	AA352154	AA352154 EST60129
39	14.8	77.9	451	33	AA436631	AA436631 zw55f03.s
C 40	14.8	77.9	453	33	AA442113	AA442113 zw55f03.r
C 41	14.8	77.9	586	35	U92781	U92781 EST MESF142
C 42	14.8	77.9	333	38	AA741801	AA741801 LMLv39p3/
C 43	14.8	77.9	563	39	AA886053	AA886053 ny44c07.s
C 44	14.8	77.9	580	39	C84819	C84819 C84819 Dict
C 45	14.8	77.9	540	51	A1708319	A1708319 atC0401.x

#### ALIGNMENTS

RESULT 1  
T64274/c 371 bp mRNA EST 17-FEB-1995  
LOCUS yc09b09.r1 Striatogene Lung (937210) Homo sapiens cDNA clone  
DEFINITION IMAGE:80153 5', mRNA sequence.  
ACCESSION T64274  
NID g668139  
VERSION T64274.1 GI:668139

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS 1 (bases 1 to 371)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Maris, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R., and Marra, M.  
TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
Insert Size: 57  
High quality sequence stops: 205 Source: IMAGE Consortium, LNL. This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 57 Std Error: 0.00  
Seq primer: M13Rpl  
High quality sequence stop: 205.  
Location/Qualifiers  
1. 371  
/organism="Homo sapiens"  
/db\_xref="GDB:483770"  
/db\_xref="taxon:9606"  
/map="14"  
/clone="IMAGE:80153"  
/clone\_lib="Stratagene lung (#937210)"  
/sex="male"  
/dev\_stage="72 years"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: lung; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. normal lung. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."  
BASE COUNT 108 a 86 c 64 g 108 t 5 others  
ORIGIN

Query Match 86.3%; Score 16.4; DB 21; Length 371;  
Best Local Similarity 94.4%; Pred. No. 62;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatgattagcagaggt 18  
||||| |||||||  
Db 62 GATGAGTAGCAGAGGCT 45

RESULT 2  
R59018/c R59018 499 bp mRNA EST 24-MAY-1995  
LOCUS y996c08.r1 Soares infant brain INTB Homo sapiens cDNA clone  
DEFINITION IMAGE:41340 5', mRNA sequence.  
ACCESSION R59018  
NID 9828713  
VERSION R59018.1 GI:829713  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS 1 (bases 1 to 499)

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Wellamson, A., Wohlmann, P., and Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On May 8, 1995 this sequence version replaced gi:800058.

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
Insert Size: 1290  
High quality sequence stops: 388 Source: IMAGE Consortium, LNL. This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1290 Std Error: 0.00  
Seq primer: M13Rpl  
High quality sequence stop: 388.  
Location/Qualifiers  
1. 499  
/organism="Homo sapiens"  
/db\_xref="GDB:413881"  
/db\_xref="taxon:9606"  
/clone="IMAGE:41340"  
/clone\_lib="Soares infant brain INTB"  
/sex="female"  
/dev\_stage="73 days post natal"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: whole brain; Vector: Lambda B; Site\_1: Not I; Site\_2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACGAGAGATTCGCGCCGACAGATTTTCTTTTCTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lambda B vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 145 a 107 c 95 g 145 t 7 others  
ORIGIN

Query Match 86.3%; Score 16.4; DB 22; Length 499;  
Best Local Similarity 94.4%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatgattagcagaggt 18  
||||| |||||||  
Db 301 GATGAGTAGCAGAGGCT 284

RESULT 3  
H61749/c H61749 374 bp mRNA EST 06-OCT-1995  
LOCUS yu39a09.r1 Soares ovary tumor NBHOT Homo sapiens cDNA clone  
DEFINITION IMAGE:236152 5', mRNA sequence.  
ACCESSION H61749  
NID 91014581  
VERSION H61749.1 GI:1014581  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS 1 (bases 1 to 374)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Maris, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Maria, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
On Sep 21, 1992 this sequence version replaced g1:276271.

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
Insert Size: 1177  
High quality sequence stops: 289  
Source: IMAGE Consortium, LNLN  
This clone is available royalty-free through LNLN; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1177 Std Error: 0.00  
Seq primer: M13RPI  
High quality sequence stop: 289.  
Location/Qualifiers

## FEATURES

source

1. 374  
/organism="Homo sapiens"  
/db\_xref="GDB:3863071"  
/db\_xref="taxon:9606"  
/clone="IMAGE:236152"  
/clone\_1lb="Soares ovary tumor NBHOT"  
/sex="Female"  
/tissue\_type="ovarian tumor"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: ovary; Vector: p773D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTACCAATCTGAGTGGAGCGCCGCCCAATTTTTTTTTTTTTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified p773 vector  
(Pharmacia). Library constructed by Bento Soares and  
M. Fatima Bonaldo."

BASE COUNT 123 a 88 c 49 g 113 t 1 others

Query Match 86.3%: Score 16.4; DB 24; Length 374;  
Best Local Similarity 94.4%: Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatgattagcagaggt 18  
||||| |||||||  
Db 114 GATGAGTACGACGAGGT 97

RESULT 4  
AA393708 239 bp mRNA EST 12-AUG-1997  
LOCUS z71910.f1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:727842  
DEFINITION 5', mRNA sequence.  
ACCESSION AA393708  
NID 92046677  
VERSION AA393708.1 GI:2046677  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

## REFERENCE

AUTHORS

1 (bases 1 to 239)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Kucaba, T., Lacy, M., Le, N., Lennon, G., Maria, M., Martin, J.,  
Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
White, Y., Wylie, T., Waterston, R. and Wilson, R.  
TITLE  
JOURNAL  
Unpublished (1997)

COMMENT On May 18, 1995 this sequence version replaced g1:810960.

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
This clone is available royalty-free through LNLN; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 924 Std Error: 0.00  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 226.  
Location/Qualifiers

## FEATURES

source

1. 239  
/organism="Homo sapiens"  
/db\_xref="GDB:5924753"  
/db\_xref="taxon:9606"  
/map="6"  
/clone="IMAGE:727842"  
/clone\_1lb="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: p773D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5'  
TGTACCAATCTGAGTGGAGCGCCGCCCAATTTTTTTTTTTTTTTT 3']  
Double-stranded cDNA was ligated to Eco RI adapters  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified p773 vector. Library  
went through one round of normalization to Cot5, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 80 a 56 c 37 g 66 t

Query Match 86.3%: Score 16.4; DB 33; Length 239;  
Best Local Similarity 94.4%: Pred. No. 57;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gatgattagcagaggt 18  
||||| |||||||  
Db 46 GATGAGTACGACGAGGT 29

RESULT 5  
H01442 309 bp mRNA EST 19-JUN-1995  
LOCUS y199h09.s1 Soares\_placenta\_NB2HP Homo sapiens cDNA clone  
DEFINITION IMAGE:147423.3, mRNA sequence.  
ACCESSION H01442  
NID 9864375  
VERSION H01442.1 GI:864375  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

## REFERENCE

AUTHORS

1 (bases 1 to 309)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Parsons, D., Rifkin, L., Roehling, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
TITLE  
JOURNAL  
Unpublished (1995)  
On May 9, 1995 this sequence version replaced g1:803733.

CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
Insert Size: 787

High quality sequence stops: 286

Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert length: 787 Std Error: 0.00  
Seq primer: Promega -21m13

High quality sequence stop: 286.

#### FEATURES

source

Location/Qualifiers  
1..309  
/organism="Homo sapiens"  
/db\_xref="GDB:559076"  
/db\_xref="taxon:9606"  
/clone="IMAGE:147425"  
/clone\_lib="Soares placenta Nb2HP"  
/sex="female"  
/dev\_stage="placenta obtained at birth (full term)"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I...oligo(dT) primer [5'  
AAGTGAAGATTCGCGCGCGAGGAAATTTTCTTTTCTTTT 3']  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 117 a 45 c 55 g 86 t 6 others

#### ORIGIN

##### Query Match

Best Local Similarity 83.2%; Score 15.8; DB 22; Length 309;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gatgattacagagaggtt 19

Db 290 GATGATTGTAGAGGCTT 308

RESULT 6  
LOCUS AI353711 420 bp mRNA EST 04-JAN-1999

DEFINITION zeb0888, seg.F zebrafish Embryonic Heart cDNA library Danio rerio

ACCESSION CDNA 5', mRNA sequence.

VERSION AI353711  
NID 94093864  
AI353711.1 GI:4093864

KEYWORDS EST.

SOURCE zebrafish.

ORGANISM

Danio rerio

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
Cyprinidae; Cyprinidae; Rasbora; Danio.  
1 (bases 1 to 420)

REFERENCE  
AUTHORS Ton, C., Mabry, J.D., Dempsey, A.A., Hwang, D.M., Fishman, M.C. and  
Liew, C.C.

TITLE Identification and Characterization of Expressed Sequence Tags from  
an Embryonic Zebrafish Heart cDNA Library

JOURNAL Unpublished (1999)

COMMENT On Jan 14, 1998 this sequence version replaced gi:1798250.

Contact: Liew CC  
Department of Laboratory Medicine and Pathobiology  
University of Toronto

Banting Institute, 100 College St., Toronto, Ontario, M5G1L5

Tel: 4169788758

Fax: 4169785650  
Email: liewcc@utcc.utoronto.ca  
PCR Primers

FORWARD: 5' GCCAAGCTCGAATTAACCTCACTAAGG 3'  
BACKWARD: 5' CCAGTGAATGTATACGACTCACTATAGGCG 3'  
Seq primer: 5' GAATTAACCTCACTAAGG 3'

Location/Qualifiers  
1..420

/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/map="6"  
/clone\_lib="zebrafish Embryonic Heart cDNA Library"  
/dev\_stage="embryonic day 3 post-fertilization"  
/lab\_host="E. coli XL1-Blue mrf"  
/note="Organ: heart; Vector: Lambda ZAP Express; Site\_1:  
EcoRI; Site\_2: XhoI; mRNA was purified from embryonic  
zebrafish hearts (3 day post-fertilization). cDNA was  
synthesized using a XhoI-Oligo dT adaptor-primer. EcoRI  
adaptors were ligated, followed by digestion with XhoI,  
for directional cloning into pre-digested lambda ZAP  
Express vector."

BASE COUNT 99 a 95 c 91 g 135 t

#### ORIGIN

##### Query Match

Best Local Similarity 83.2%; Score 15.8; DB 45; Length 420;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gatgattacagagaggtt 19

Db 52 GATGCTTACGACGACGCTT 70

RESULT 7  
LOCUS T09332 407 bp mRNA EST 03-AUG-1993

DEFINITION ESR07225 Infant Brain, Bento Soares Homo sapiens cDNA clone HIBBS49

ACCESSION 5' end, mRNA sequence.

VERSION T09332  
NID 9390360  
T09332.1 GI:390360

KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C. and Venter, J.C.

TITLE 1 (bases 1 to 407)  
Rapid cDNA sequencing (expressed sequence tags) from a  
directionally cloned human Infant Brain cDNA library

JOURNAL Nature Genet. 4, 373-380 (1993)

COMMENT 94004965

Contact: Adams, MD  
The Institute for Genomic Research  
932 Clopper Road, Gaithersburg, MD 20878

Tel: 3018699056

Fax: 3018699423

Email: mdadams@tigr.org

Seq primer: M13 Reverse.

#### FEATURES

source

Location/Qualifiers  
1..407  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):85787"  
/db\_xref="taxon:9606"  
/clone="HIBBS49"

BASE COUNT 115 a 83 c 98 g 105 t 6 others

#### ORIGIN

##### Query Match

Best Local Similarity 81.1%; Score 15.4; DB 20; Length 407;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;





sequence.  
 R58460  
 R582818  
 R58460.1 GI:828518  
 EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 267)  
 Hwang D.M., Fung Y.W., Wang R.X., Laurensen C.M., Ng S.H., Lam W.Y., Tsui K.W., Fung K.P., Maye M., Lee C.Y. and Liew C.C. Analysis of expressed sequence tags from a fetal human heart cDNA library  
 Genomics 30 (2), 293-298 (1995)  
 JOURNAL 96163883  
 MEDLINE  
 COMMENT On May 8, 1995 this sequence version replaced gi:799893.  
 Other ESTs: gb T09332  
 Contact: Liew CC  
 Department of Laboratory Medicine and Pathobiology  
 University of Toronto  
 Banting Institute, 100 College St., Toronto, Ontario, M5G1L5  
 Tel: 4169788758  
 Fax: 4169785650  
 Email: liewcc@utcc.utoronto.ca  
 Insert length: 1016 Std Error: 0.00  
 Seq primer: GGTGGCAGACTCTCGAGCC  
 High quality sequence stop: 180.  
 Location/Qualifiers  
 1..267  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="G3742"  
 /clone\_lib="Fetal heart"  
 /lab\_host="E. coli Y1090"  
 /note="Vector: Lambda gt22; Site\_1: NotI; Site\_2: SalI; mRNA was purified from human fetal hearts (10-12 weeks). cDNA was constructed using a NotI-0.10g dt adaptor primer. Salt adaptors were ligated, followed by digestion with NotI, for direction cloning into predigested lambda gt22. Method is described in J. Mol. Cell. Cardiol. (1994) 26, 1329-1333")  
 BASE COUNT 68 a 57 c 71 g 71 t  
 ORIGIN  
 Query Match 81.1%; Score 15.4; DB 22; Length 267;  
 Best Local Similarity 94.1%; Pred. No. 1.9e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 gatgattagcagaggt 17  
 ||||| ||||| |||||  
 Db 68 GATGCTTACGACAGGCG 84  
 RESULT 11  
 LOCUS AA002971/c 446 bp mRNA EST 19-JUL-1996  
 DEFINITION mg50e04.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA  
 clone IMAGE:427230 5', mRNA sequence.  
 ACCESSION AA002971  
 NID 91446410  
 VERSION AA002971.1 GI:1446410  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 446)  
 Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T., Geisels S., Kucaba T., Lacy M., Le M., Martin J., Morris M., Scheinberg K., Steptoe M., Yan F., Underwood K., Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R. and

Waterston R.  
 The WashU-HMI Mouse EST Project  
 Unpublished (1996)  
 COMMENT  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:261782  
 Seq primer: ETPprimer  
 High quality sequence stop: 356.  
 Location/Qualifiers  
 1..446  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:427230"  
 /clone\_lib="Soares mouse embryo NBME13.5 14.5"  
 /sex="unknown"  
 /tissue\_type="embryo"  
 /dev\_stage="13.5-14.5dpc total fetus"  
 /lab\_host="DHI08"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - 0.10g(dt) primer (5' TGTTCACATCTGAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTT T 3'), on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos (total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."  
 BASE COUNT 118 a 117 c 110 g 101 t  
 ORIGIN  
 Query Match 81.1%; Score 15.4; DB 27; Length 446;  
 Best Local Similarity 94.1%; Pred. No. 2.2e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 atgattagcagaggt 18  
 ||||| ||||| |||||  
 Db 295 ATGATGAGCAGAGGCT 279  
 RESULT 12  
 LOCUS AA090488 278 bp mRNA EST 24-OCT-1996  
 DEFINITION y0582.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA 5', mRNA sequence.  
 ACCESSION AA090488  
 NID 91635072  
 VERSION AA090488.1 GI:1635072  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 278)  
 Liew C.C.  
 cDNAs from fetal heart (1996)  
 JOURNAL Unpublished (1996)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1397609.  
 Contact: Liew CC  
 Department of Laboratory Medicine and Pathobiology

University of Toronto  
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5  
Tel: 4169788758  
Fax: 4169785650  
Email: liwcc@utcc.utoronto.ca  
PCR Primers  
FORWARD: 5' GCCAAGCTCGAATTACCTCTACTAAGG 3'  
BACKWARD: 5' CCAGTAATGTAATAGCAGCTACTATAGGCG 3'  
Seq primer: 5' GAATTACCTCTACTAAGG 3'  
Location/Qualifiers  
1. 278  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Human fetal heart, Lambda ZAP Express"  
/lab\_host="E. coli XL1-Blue"  
/note="Vector: Lambda ZAP Express; Site\_1: EcoRI; Site\_2:  
XhoI; mRNA was purified from human fetal hearts (8-10  
weeks). cDNA was synthesized using a XhoI-Oligo dT  
adaptor-primer. EcoRI adaptors were ligated, followed by  
digestion with XhoI, for directional cloning into  
predigested lambda ZAP Express."

## BASE COUNT

80 a 37 c 58 g 103 t

## ORIGIN

Query Match 81.1%; Score 15.4; DB 28; Length 278;  
Best Local Similarity 94.1%; Pred. No. 2e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tggattagcagaggtt 19  
|||||  
Db 202 TGGAAATGACAGAGGTT 218

RESULT 13  
AA241858/c  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AA241858 560 bp mRNA EST 07-MAR-1997  
mx21c11.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:680852  
5', mRNA sequence.  
AA241858  
G1872830  
AA241858.1 G1:1872830  
EST.  
house mouse.  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 560)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
On May 8, 1995 this sequence version replaced g1:801312.

## FEATURES

source  
1. 560  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
Contact: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 488.  
Location/Qualifiers  
1. 560  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

/clone="IMAGE:680852"  
/clone\_lib="Soares mouse NML"  
/issue\_type="Liver"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTCACATCTGAGTGGAGCGCCGCGAATCTTTTCTTTTCTTTT 3'];  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
constructed and normalized by Bento Soares and M. Fatima  
Bonaldo."

BASE COUNT 170 a 130 c 137 g 123 t  
ORIGIN  
Query Match 81.1%; Score 15.4; DB 30; Length 560;  
Best Local Similarity 94.1%; Pred. No. 2.3e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgattagcagaggtt 18  
|||||  
Db 199 ATGATGACGAGAGGT 183

RESULT 14  
AA244838/c  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AA244838 342 bp mRNA EST 10-MAR-1997  
mx13a09.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:680056  
5', mRNA sequence.  
AA244838  
G1875565  
AA244838.1 G1:1875565  
EST.  
house mouse.  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 342)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
On Jan 25, 1995 this sequence version replaced g1:637877.

CONTACT: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 288.  
Location/Qualifiers  
1. 342  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:680056"  
/clone\_lib="Soares mouse NML"  
/issue\_type="Liver"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTCACATCTGAGTGGAGCGCCGCGAATCTTTTCTTTTCTTTT 3'];

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

Query Match 81.1%; Score 15.4; DB 30; Length 342;  
Best Local Similarity 94.1%; Pred. No. 2.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgattagcagaggt 18  
||||| |||||  
DB 319 ATGATGACGACAGCGT 303

RESULT 15  
AA261440/C 188 bp mRNA EST 18-MAR-1997  
LOCUS m287f02.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:720411  
DEFINITION 5' mRNA sequence.  
ACCESSION AA261440  
NID 91895172  
VERSION AA261440.1 GI:1895172  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 188)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1395029.

CONTACT: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:445907  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 161.

# FEATURES

## SOURCE

1. 188  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:720411"  
/clone\_lib="Soares mouse NML"  
/tissue\_type="liver"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer (5'  
TGTTCACATCTGAGCGGCGCCGCGATCTTTTCTTTTCTTTT 3');  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."

BASE COUNT  
ORIGIN

53 a 43 c 49 g 43 t

Query Match 81.1%; Score 15.4; DB 30; Length 188;  
Best Local Similarity 94.1%; Pred. No. 1.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atgattagcagaggt 18  
||||| |||||  
DB 42 ATGATGACGACAGCGT 26

Search completed: August 18, 1999, 17:04:46  
Job time: 2792 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 17:17:19 ; Search time 543.6 Seconds

(without alignments)  
122.860 Million cell updates/sec

Title: US-09-004-395-5

Perfect score: 21

Sequence: 1 tgggataaataatgagcgct 21

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_dal:\*  
2: gb\_baz:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pac:\*  
6: gb\_ph:\*  
7: gb\_pil:\*  
8: gb\_pil2:\*  
9: gb\_pil3:\*  
10: gb\_pil4:\*  
11: gb\_pil5:\*  
12: gb\_pil6:\*  
13: gb\_pil7:\*  
14: gb\_pil8:\*  
15: gb\_pil9:\*  
16: gb\_pil10:\*  
17: gb\_pil11:\*  
18: gb\_pil12:\*  
19: gb\_pil13:\*  
20: gb\_pil14:\*  
21: gb\_pil15:\*  
22: gb\_pil16:\*  
23: gb\_pil17:\*  
24: gb\_pil18:\*  
25: gb\_pil19:\*  
26: gb\_pil20:\*  
27: gb\_pil21:\*  
28: gb\_pil22:\*  
29: gb\_pil23:\*  
30: gb\_pil24:\*  
31: gb\_pil25:\*  
32: gb\_pil26:\*  
33: gb\_pil27:\*  
34: gb\_pil28:\*  
35: gb\_pil29:\*  
36: gb\_pil30:\*  
37: gb\_pil31:\*  
38: gb\_pil32:\*  
39: gb\_pil33:\*  
40: gb\_pil34:\*  
41: gb\_pil35:\*  
42: gb\_pil36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	21	100.0	13271	2	AE001168	AE001168 Borrelia

2	21	100.0	1655	2	BBU62900	U62900 Borrelia bu
3	18.4	87.6	1420	37	AF080564	AF080564 Anopheles
4	17.4	82.9	183057	34	AF127019	AF127019 Homo sapi
5	17	81.0	39602	10	HSAC000373	AC000373 Homo sapi
6	17	81.0	130336	34	AC006346	AC006346 Homo sapi
7	17	81.0	18633	36	CELR24C12	U41037 Caenorhabd
8	16.8	80.0	1528	1	BS16086	AU05086 Bacillus
9	16.8	80.0	55721	8	TIF15	AC004393 Arabidops
10	16.8	80.0	144514	34	HS1018D12	AL031650 Homo sapi
11	16.8	80.0	211912	35	AC007437	AC007437 Homo sapi
12	16.8	80.0	171236	42	AC007656	AC007656 Homo sapi
13	16.4	78.1	129440	11	AC004806	AC004806 Homo sapi
14	16.4	78.1	46127	11	AC005347	AC005347 Homo sapi
15	16.4	78.1	19259	17	CITY18420	Y18420 Citrus tils
16	16.2	77.1	338534	1	ECOMW93	Y10168 B. thuringie
17	16.2	77.1	13889	2	AE000487	U14003 Escherichia
18	16.2	77.1	82010	7	AB023042	AB023042 Escherich
19	16.2	77.1	41275	7	SPBC1734	AL031856 S. pombe c
20	16.2	77.1	37342	7	SPBC1734	AL031854 S. pombe c
21	16.2	77.1	107600	8	ATF16J13	AL049638 Arabidops
22	16.2	77.1	90824	8	ATT28M18	AL078606 Arabidops
23	16.2	77.1	78025	9	HS780M13	AL022146 Human DNA
24	16.2	77.1	93431	11	HU095737	U93737 Human Chrom
25	16.2	77.1	146070	34	AC006595	AC006595 Homo sapi
26	16.2	77.1	204104	34	CNS0000A	AL049828 Homo sapi
27	16.2	77.1	4675	36	DDU20608	U20608 Dictyostell
28	16.2	77.1	207319	35	AC006517	AC006517 Homo sapi
29	16	76.2	207319	35	AC006517	AC006517 Homo sapi
30	16	76.2	2291870	35	AC006927	AC006927 Homo sapi
31	15.8	75.2	172000	1	AP000005	AP000005 Pyrococcu
32	15.8	75.2	3246	1	HPKATA	Z70679 H. pylori ka
33	15.8	75.2	135638	1	SYCSLRG	D64005 Synechocyst
34	15.8	75.2	9854	2	AE000597	AE000597 Helicobac
35	15.8	75.2	10311	2	AE001510	AE001510 Helicobac
36	15.8	75.2	1769	2	HPU67458	U67458 Helicobacte
37	15.8	75.2	735	3	AF053497	AF053497 Equus cab
38	15.8	75.2	778	3	AF068057	AF068057 Oryctolag
39	15.8	75.2	1068	5	A21606	A21606 human FC ep
40	15.8	75.2	1198	5	109686	109686 Sequence 3
41	15.8	75.2	83450	7	AB009054	AB009054 Arabidops
42	15.8	75.2	67720	7	AB019224	AB019224 Arabidops
43	15.8	75.2	40214	8	YSC18003	U17243 Saccharomyc
44	15.8	75.2	7659	9	HUMIGERA	L14075 Homo sapien
45	15.8	75.2	42305	35	AC007826	AC007826 Drosophll

## ALIGNMENTS

RESULT	1	LOCUS	AE001168	DEFINITION	Borrelia burgdorferi	ACCESSION	AE001168	VERSION	AE001168.1	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
LOCUS	AE001168	13271 bp	DNA	BCT	15-DEC-1997	LOCUS	AE001168	13271 bp	DNA	BCT	15-DEC-1997	DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
ACCESSION	AE001168	AE000783				ACCESSION	AE001168	AE000783				DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
VERSION	AE001168.1	GI:2688598				VERSION	AE001168.1	GI:2688598				DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
KEYWORDS						KEYWORDS						DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
ORGANISM						ORGANISM						DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
REFERENCE	1	(bases 1 to 13271)				REFERENCE	1	(bases 1 to 13271)				DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
AUTHORS	Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,					AUTHORS	Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,					DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
	Watt,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D.,						Watt,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D.,					DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
	Richardson,D., Peterson,J., Kierlavage,A.R., Quackenbush,J.,						Richardson,D., Peterson,J., Kierlavage,A.R., Quackenbush,J.,					DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
	Salzberg,S., Hanson,M., Van-Vugt,R., Palmer,N., Adams,M.D.,						Salzberg,S., Hanson,M., Van-Vugt,R., Palmer,N., Adams,M.D.,					DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
	Gocayne,J.D., Weidman,J., Utterback,T., Matthey,L., McDonald,L.,						Gocayne,J.D., Weidman,J., Utterback,T., Matthey,L., McDonald,L.,					DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
	Artlich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,						Artlich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,					DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
	Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.						Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.					DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
	Genomic sequence of a Lyme disease spirochete, Borrelia						Genomic sequence of a Lyme disease spirochete, Borrelia					DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
	burgdorferi						burgdorferi					DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		
	Nature 390 (6660), 580-586 (1997)						Nature 390 (6660), 580-586 (1997)					DEFINITION	Borrelia burgdorferi (section 54 of 70) of the complete genome.		

MEDLINE  
REFERENCE  
AUTHORS  
98065943  
2 (bases 1 to 13271)  
Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,  
Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,  
Gwin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,  
Richardson,D., Peterson,J., Kellavagge,A.R., Quackenbush,J.,  
Salzberg,S., Hanson,M., Van Vugt,R., Palmer,N., Adams,M.D.,  
Gocayne,J.D., Weidman,J., Utecherback,T., Wathey,L., McDonald,L.,  
Artach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,  
Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.  
TITLE  
JOURNAL  
Submitted (12-DEC-1997) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
source  
1..13271  
/organism="Borrelia burgdorferi"  
/db\_xref="taxon:139"  
66..1100  
/gene="BB0668"  
66..1100  
/note="BB0668"  
/note="similar to GP:1575447 percent identity: 98.83;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="flagellar filament outer layer protein (flaa)"  
/protein\_id="AAC67025.1"  
/db\_xref="PID:g2688608"  
/db\_xref="GI:2688608"  
/translation="MYMKRKAASLIFLSTVLAQETDGAESKRAPEGLVLD  
AELADPSRLDITNVDYVSGAVIKPDMVVDGINSVSLTPASALAVK  
NSVAPAVVSKSRAGDTILGVRILPSPYSQSSAMIPPKIPFSGESGNOFLG  
GLIDIKTKKEIKVSVISGYEIDLEVEFEDMNGMAYSMGLTKFKGADLVMSNP  
YIPNISRIIKDQVNPYPLASMKRFKARVSKSHSKQNFIFVYKDLRYDLSY  
SIDSDISESVKVVETSGTESLRLLKAETFKVRLKREKISMESGNSQNVKIES  
EXPRESSPKN"  
1168..3762  
/gene="BB0669"  
1168..3762  
/note="BB0669"  
/note="similar to PID:1113815 SP:044737 GB:AE000783  
percent identity: 99.31; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis histidine kinase (cheA-2)"  
/protein\_id="AAC67024.1"  
/db\_xref="PID:g2688607"  
/db\_xref="GI:2688607"  
/translation="MEILDLENEELLGVFEBAQNLVDILEENINSLDEDDPNNSDTID  
EIFRAHRLKGGSSADMMELSDFTIIVEDVDARIGKVININDLVLLSSLDVIR  
EMALRIDKRYLINDISDLKSKLOFLVDQTFIKRFGNSIKNNFCLESDLEIR  
EGLIGOKVLRISVFNSSNGSEVENSGLKINILKNGSVLHTIPKYEOLIEDFK  
RDYLYIYDISGVKKSLSLNLISYLVDEPNVKELKKLADDEIKVDSDSNVLA  
DNPFTBEDISDLLEVENQKLFKRLDVKNPMTISGLDMLQALSLGKIFRSIP  
DSSELLADFEFVLYIYLISNTSESIARKNLPVVSFEIKNVNLSLKVRLKED  
DEAPFENKKNKNSPIVNLIRIDSKIDYLLINSEAVYISKSSYNOINSMTILFY  
NNFYVDQESFORNLIDKLVKQDGLTLEDESHSNLSFMKALDISELRL  
NSFPLIIONKMTSGRLSRLITDLSHSLTKMLPISNFSRTRVYRDLSSKLLIV  
NKMGEFEELDKSVYDLDVPLMLCYRSMHSGLETVEERYKRGSKRAGTITLAKN  
EGNVSIEIEDGIGIDPKVIRKLEKGTIKEDATISDFELINLIFAGFSTAQVY  
DLSGRGVGLDVVRSKTEKLNGTILVESEIGLQTIKIKLPLVLVIOGLVSGSEYI  
VPLNNVLETHRIETEDHIDKLENYHEVNLREDEVSLRLDKLFNTRDSDLEFELI  
VNTSNMKIAIVVDSILGDEDEVKPKINDKFSAGIYGATTLGNGKVVLLIDVERKL  
DLQDKRKE"  
gene  
3769..5169  
/gene="BB0670"  
3769..5169  
/note="BB0670"  
/note="similar to GP:1881576 percent identity: 98.93;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11

/product="purine-binding chemotaxis protein (cheW-3)"  
/protein\_id="AAC67023.1"  
/db\_xref="PID:g2688606"  
/db\_xref="GI:2688606"  
/translation="MOIRKEGPKRIIDDKNSKSLNPFDKVVSFELSGHYLVDM  
QVKERKSSNTIYPNAKKYVAGDNLRLGELIPIIDIRIMNLEFNKKDIDIVLKN  
EDLIGIVVDINVSIDSLLDPPVLSODSLINTIKGVYNEKTLILVNFKI  
FNVEEELKLPQNFVEKSDVSDCNLDLLENSKNDVSKVSKNDGNENSTLNN  
TAANLEIKRNLKYSFNASLVNDVLEKGVFNVDYDLPDGSFLNFEYSKSG  
NMGADCLKEERKNELVESRLSEFNNNLSIFNVLIEIGSGKETAALNALSEYVKEP  
KLTAINDLSKVSFETSRVPSSEISGISEYRSNFSPOSPVYKFEKSEIISNVFEYS  
DALSDLPDNGMFLKDYLCFLDSKOIILNITASKAKGALLVLDNDELKNNDV  
FIKKESTYFENLYRKV"  
gene  
5181..5666  
/gene="BB0671"  
5181..5666  
/note="BB0671"  
/note="similar to GP:1408273 percent identity: 98.76;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis operon protein (chex)"  
/protein\_id="AAC67022.1"  
/db\_xref="PID:g2688605"  
/db\_xref="GI:2688605"  
/translation="MRIDYIEPFLDAASVLRDMLVNETEMGRKGLSKINOKIGVS  
VIVGLAGSEGSIIIDMDIETALVNASLNEEVDDEDETKEMVAATLVEVNIIA  
GNFTYTHAKGFVDIIPPAFVYGENKKSIGSEALIVPSPLDGKIEVNIARER  
V"  
5706..6146  
/gene="BB0672"  
5706..6146  
/note="BB0672"  
/note="similar to GP:1408274 percent identity: 97.95;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="chemotaxis response regulator (chey-3)"  
/protein\_id="AAC67021.1"  
/db\_xref="PID:g2688604"  
/db\_xref="GI:2688604"  
/translation="MIQKTTIAADSSSKPRGINDTGIPNVLLVYDVSFTVRLTOI  
FTSBGFNIDTPADGEEAVIKNNHYPIDIVLTDITMPKDGITICUSNIEFKNAR  
VIMSALKEQLVDCILGAKTIVKPLDRAKVLQVMSVFK"  
complement(6187..6702)  
/gene="BB0673"  
complement(6187..6702)  
/note="BB0673"  
/note="similar to GP:1408275 percent identity: 97.56;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAC67020.1"  
/db\_xref="PID:g2688603"  
/db\_xref="GI:2688603"  
/translation="MKIQIILMLALDLPUNARLDISIKRADEIKKXSYNLLI  
EKEVYINPTESEIKNIKLEHVPVKSIMLKNTVSLNSYKANKYALLOSELIDK  
FLKIKPIKIKINGINIFKSHSLYIKKGGYKLELTIENNAEPLKIFNLNTYFLNLNDK  
ISNKKIFPRR"  
complement(6699..7745)  
/gene="BB0674"  
complement(6699..7745)  
/note="BB0674"  
/note="hypothetical protein; identified by Glimmer;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="B. burgdorferi predicted coding region BB0674"  
/protein\_id="AAC67027.1"  
/db\_xref="PID:g2688610"  
/db\_xref="GI:2688610"  
/translation="MYKKFSIFLKAIIIFSIFELLIELSLITLFPYKIRPALIFGR

```

gene
CDS
    complement(7739..8602)
    /gene="BB0675"
    complement(7739..8602)
    /gene="BB0675"
    /note="hypothetical protein; identified by Glimmer:
    putative"
    /codon_start=1
    /transl_table=11
    /product="B. burgdorferi predicted coding region BB0675"
    /protein_id="AAC67026.1"
    /db_xref="PID:92688609"
    /db_xref="GI:2688609"
    /translation="MNTKRPYPYLHYLFNHESIKLSAIEKEIEILNYLKENKTIA
    TEIKNDPSEIKDLQYVKDKDILIPFVSGEIAIDENFVILFKSLKRNLDNLI
    FNEVNVNSLKEKPEYKNTISNGYITPYIKLPEGKSGVYIQLQENRAYSDDIIV
    KNYIILLILLYKPKYKCEKGISETLTNINISTKASINDPFDLAKTALISSEFK
    YETLDTMSPISTLILAIISARAKIPKTKNNPVKGIIDESWFSIKOSGSRYSRIRIK
    ELSEIAKNNKW"
    complement(8599..9261)
    /gene="BB0676"
    complement(8599..9261)
    /gene="BB0676"
    /note="similar to GB:IL4203 SP:P44755 PID:1004013
    PID:1223502 PID:1204815 percent identity: 25.58;
    identified by sequence similarity; putative"
    /codon_start=1

```

```

Query Match          100.0% ; Score 21 ; DB 2 ; Length 13271;
Best Local Similarity 100.0% ; Pred. No. 1.9 ;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tgggataataatctgagcgt 21
    |||||
Db 311 TGGGATAAATAATTGGAGCGT 331

```

```

RESULT      2
LOCUS       BBu62900
DEFINITION  BBu62900      1655 bp      DNA      BCT      15-JAN-1997
            Borrelia burgdorferi flagellar filament outsheath protein (flaA
            gene, complete cds), and chemotaxis histidine kinase (cheA) gene
            partial cds.
ACCESSION   U62900
VERSION     91575445
KEYWORDS    U62900.1  GI:1575445
SOURCE      Lyme disease spirochete.
ORGANISM   Borrelia burgdorferi
            Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
            burgdorferi group.
            1 (bases 1 to 1655)
            Ge.Y. and Charon,M.W.
            An unexpected FlaA homolog is present and expressed in Borrelia
            burgdorferi
            J. Bacteriol. 179 (2), 552-556 (1997)
            97144545
            2 (bases 1 to 1655)
            Ge.Y.
            Direct Submission
            Submitted (03-JUL-1996) Yigong Ge, West Virginia University,
            Microbiology, HSCN, Morgantown, WV 26506, USA
            Location/Qualifiers
            1. 1655
               /organism="Borrelia burgdorferi"
               /strain="212"
               /db_xref="taxon:139"
               /clone="fwi and pw2"
FEATURES
Source

```

```

CDS
gene
>1..395
/function="unknown"
/note="orfA"
/codon_start=3
/transl_table=1
/protein_id="AAC44769.1"
/db_xref="PID:91575446"
/db_xref="GI:1575446"
/transl_table="DPLESKRYGFHSHSGCVANLRLYLRDQNLKNDLVLKPFLLG
SNEGSPFELLSDSRELFYSKNGVYVNLRFREYDSDPNNGDYNESFDVKNLPLAMSJ
IKTMFNFPGIKRKRIILVEGCILKEQS"
473..1498
/gene="flaA"
473..1498
/gene="flaA"
/function="putative flagellar filament outsheath protein"
/codon_start=1
/transl_table=1
/product="flaA protein"
/protein_id="AAC44770.1"
/db_xref="PID:91575447"
/db_xref="GI:1573447"
/transl_table="MKRAKSIILFFLSTVLFAOETDGLAEGSKRAEGBELVDLPAEI
AADPSTILDLTINYDVIYSGASGIVPKEDAVVLDGINNWSVLLTPSARLQAYVKNSS
VAPVAEESKSRADITILGVRLVFPYSQSASAMIMPFKIPFYSGESGNOLFCKGKGL
DNIKMRKIKTYSYSLAIEIDLEVLFEEDMNEEYAYSGTILKFGMALILSNSTYIFP
NNSIILIDQVNPVLASSKRMFKAFRPFKSHSSKSVKKEFIYVDLRYLVKLSVSDIS
SDISESYFKYETISGTESLRKRAHEFKRYVLKREKISIAESFQNFVEKIESERR
EESSPKRN"
1566..1655
/gene="cheA"
1566..>1655
CDS

```

	BASE COUNT	560 a	182 c	308 g	605 t	
	ORIGIN					
	Query Match	100.0%; Score 21; DB 2; Length 1655;				
	Best Local Similarity	100.0%; Pred. NO. 2.3;				
	Matches	21; Conservative	0;	Mismatches	0;	Indels
	Gaps	0;				
OY	1 tgggataataatcttgagcgt	21				
DB	709 TGGGATAAATAATTGGAGCGT	729				

RESULT 3  
 AF080564/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 NID  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

AF080564 1420 bp mRNA INV 16-AUG-1998  
 Anopheles gambiae Sex combs reduced homeotic protein (Scr) mRNA,  
 complete cds.  
 AF080564  
 g3420833  
 AF080564.1 GI:3420833  
 African malaria mosquito.  
 Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;  
 Culicoidae; Culicidae; Anophelinae.  
 1 (bases 1 to 1420)  
 Devenport, M. P. and Eggleston, P.  
 Direct Submission  
 Submitted (29-JUL-1998) School of Biological Sciences, University  
 of Liverpool, Donnan Laboratories, Liverpool L69 7BD, UK





```

/note="GRAIL, score = 47.000%, comment = marginal shadow"
/evidence-not_experimental
exon 10459. .10563
/note="GRAIL, score = 69.000%, comment = good"
/evidence-not_experimental
repeat_region complement(112621. .12526)
/rpt_family="L2"
/evidence-not_experimental
exon 12886. .13029
/note="GRAIL, score = 45.000%, comment = marginal"
/evidence-not_experimental
repeat_region 12987. .13152
/rpt_family="MIR"
/evidence-not_experimental
repeat_region complement(13288. .13412)
/rpt_family="L2"
/evidence-not_experimental
repeat_region 13797. .13864
/note="homology = 91.20%, score = 22, counts = 2"
/rpt_type="candem
/rpt_unit="aaactgatttaactgtaaaactgataataa
/evidence-not_experimental
15687. .15955
/rpt_family="L1MAL"
/evidence-not_experimental
exon 16465. .16618
/note="WZEF, score = 90.7%"
/evidence-not_experimental
repeat_region 17635. .17885
/rpt_family="AluDb"
/evidence-not_experimental
exon 17775. .17996
/note="WZEF, score = 93.8%"
/evidence-not_experimental
complement(118091. .18172)
/note="Genscan, score = 1.48%, comment = Internal_exon 82
bp frame: 2 phase: 1
/evidence-not_experimental
exon complement(18091. .18172)
/evidence-not_experimental
exon complement(18091. .18172)
/note="WZEF, score = 77.2%"
/evidence-not_experimental
18444. .18515
/note="WZEF, score = 74.5%"
/evidence-not_experimental
exon complement(18568. .18634)
/note="Xpound exon prediction, score = 86% (0%)"
/evidence-not_experimental
repeat_region 18754. .19058
/rpt_family="AluSq"
/evidence-not_experimental
repeat_region complement(20537. .20832)
/rpt_family="AluSx"
/evidence-not_experimental
exon 20558. .20813
/note="WZEF, score = 50.5%"
/evidence-not_experimental
repeat_region complement(21617. .21918)
/rpt_family="AluY"
/evidence-not_experimental
exon complement(21620. .21979)
/note="WZEF, score = 66.8%"
/evidence-not_experimental
22117. .22184
/note="WZEF, score = 54.4%"
/evidence-not_experimental
exon 22355. .22388
/note="GRAIL, score = 99.000%, comment = excellent shadow"
/evidence-not_experimental
complement(22868. .22962)
/note="GRAIL, score = 89.000%, comment = excellent"
/evidence-not_experimental
repeat_region complement(23081. .23195)
/rpt_family="L2"

```

```

repeat_region /evidence-not_experimental
23888. .24170
/rpt_family="AluSx"
/evidence-not_experimental
repeat_region complement(24685. .24855)
/rpt_family="MIR"
/evidence-not_experimental
exon complement(25837. .25870)
/note="Genscan, score = -2.12%, comment = Internal_exon 34
bp frame: 1 phase: 1"
/evidence-not_experimental
repeat_region 26573. .26870
/rpt_family="AluSx"
/evidence-not_experimental
exon complement(26731. .26846)
/note="WZEF, score = 89.1%"
/evidence-not_experimental
exon 27286. .27377
/note="GRAIL, score = 59.000%, comment = good"
/evidence-not_experimental
exon 28017. .28091
/note="GRAIL, score = 62.000%, comment = good"
/evidence-not_experimental
exon 28023. .28031
/note="Xpound exon prediction, score = 66% (0%)"

```

Query Match 82.9%; Score 17.4; DB 34; Length 183057;  
Best Local Similarity 94.7%; Pred. No. 68;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 tgggataaataattggagc 19
DB 36242 TGGGATACGATTAATGAGC 36224

```

RESULT 5  
HSAC000373/C  
LOCUS HSAC000373 39602 bp ms-DNA PRI 12-MAR-1997  
DEFINITION Human cosmid g1862d218, complete sequence.  
ACCESSION AC000373  
G1861570  
VERSION AC000373.1 GI:1861570  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 39602)  
Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,  
Green,P. and Olson,M.V.  
Large-scale MCD Mapping and Sequencing of Human Chromosome 7  
Unpublished (1996)  
2 (bases 1 to 39602)  
Magness,C.L.  
Direct Submission  
Submitted (12-MAR-1997) Human Genome Center, University of  
Washington, Box 352145, Seattle, WA 98195, USA  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
These values are not generally visible from the  
Genbank flat file format but are available as part  
of the ASN.1 file.  
Verification: This sequence has been verified by Multiple Complete  
Digest Mapping. Comparison of the experimentally derived map digest  
fragments with sequence-predicted fragments is given below. Small  
fragments below a variable cutoff (approximately 400-600bp) are not  
mapped and hence do not appear in the table. There are no  
significant remaining discrepancies between the experimental and  
predicted values. Uniquely ordered fragment groups are separated by  
dashed lines.

Map	Seq	Map	Seq	Map	Seq
-----	Ecort	-----	HindIII	-----	NsiI
-----	Seq	-----	Seq	-----	Seq

```

6782.98 6781.00 494.55 490.00 1775.10 1788.00
5710.44 5707.00 15143.19 15309.00 7155.88 7144.00
931.21 928.00 5350.00 5331.00 2027.07 2055.00
7869.24 7885.00 2001.88 2025.00 7163.17 7212.00
4717.89 4734.00 2977.65 2964.00 3463.65 3453.00
794.53 797.00 9310.60 9306.00 4160.82 4164.00
7176.38 7183.00 3903.12 3876.00
1651.21 1636.00 2084.45 2104.00

```

```

FEATURES
source
Location/Qualifiers
1.39602
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q31.3"
/clone="NCBGR:YMS51862"
/sub_clone="UWGC:q1862d218"
/cell_line="GM10751"
/clone_lib="E. Green Chromosome 7 YAC Resource"
5266.5336
/rpt_family="MLT2B2"
5357.5635
/rpt_family="ALU"
5741.5398
/rpt_family="MT2B2"
7554.8063
/rpt_family="L1"
complement(10571..10688)
/rpt_family="ALU"
14796.14941
/standard_name="HUMSWS3394"
/note="Genbank Accession: G13152"
16084.16489
/rpt_family="TBE1"
complement(20791..20860)
/rpt_family="ALU"
20893.21813
/rpt_family="L1"
21845.22123
/rpt_family="ALU"
complement(22879..23157)
/rpt_family="ALU"
23607.23786
/rpt_family="MIR"
27132.27219
/rpt_family="MIR"
32852.32916
/rpt_family="MIR"
34766.35067
/rpt_family="ALU"
38224.38342
/rpt_family="L1"
complement(38820..38880)
/rpt_family="MER46"
complement(39085..39163)
/rpt_family="MER46"

```

```

BASE COUNT 13122 a 7537 c 7696 g 11247 t
ORIGIN

```

```

Query Match 81.0%; Score 17; DB 10; Length 39602;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 tgggataaataattgga 17
|||||

```

```

Db 33634 TGGGATTAATAATTGCA 33618

```

```

RESULT 6
AC006346/c
LOCUS AC006346
DEFINITION Homo sapiens clone DJ0820622, WORKING DRAFT SEQUENCE, 1 unordered
pieces.
ACCESSION AC006346
NID 9419992
VERSION AC006346.1 GI:4199992
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 130336)
Waterston,R.H.
The sequence of Homo sapiens clone
2 (bases 1 to 130336)
Waterston,R.H.
Direct Submission
Submitted (11-JAN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```

COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 130336: contig of 130336 bp in length.
Location/Qualifiers
1.130336
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DJ0820622"

```

```

FEATURES
source
BASE COUNT 41771 a 25513 c 23549 g 39503 t
ORIGIN

```

```

Query Match 81.0%; Score 17; DB 34; Length 130336;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 tgggataaataattgga 17
|||||
Db 86314 TGGGATTAATAATTGCA 86298

```

```

RESULT 7
CELT24C12/c
LOCUS CELT24C12 18633 bp DNA INV 30-NOV-1995
DEFINITION Caenorhabditis elegans cosmid T24C12.
ACCESSION U41037
NID 91086755
VERSION U41037.1 GI:1086755
KEYWORDS
SOURCE Caenorhabditis elegans strain-Bristol N2.
ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 18633)
Wilson,R., Alnscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., Murray,A., Mortimore,B.,

```

```

REFERENCE
AUTHORS

```

TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

O'Callaghan, M., Parsons, J., Percy, C., Riksen, L., Roopra, A.,  
Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E.,  
Staden, R., Sulston, J., Thierly-Mieg, J., Thomas, K., Vaudin, M.,  
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,  
Wilkinson-Spratt, J. and Wohlman, P.  
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans  
Nature 368 (6466), 32-38 (1994)  
94150718  
2 (bases 1 to 18633)  
Miller, N. and Stellyes, L.  
The sequence of C. elegans cosmid T24C12  
unpublished (1995)  
3 (bases 1 to 18633)  
Waterston, R.  
Direct Submission  
Submitted (20-NOV-1995) Robert Waterston

Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RO, England  
e-mail: rwenemacode.wustl.edu and jee@sanger.ac.uk  
NEIGHBORING COSMID INFORMATION:  
The 5 cosmid is F48B9, 200 bp overlap; 3' cosmid is T14F9. Actual  
start of this cosmid is at base position 26876 of CELF48B9.

## NOTES:

Coding sequences below are predicted from computer analysis, using  
the program GeneFINDER (P. Green and L. Hillier, ms in preparation).

## FEATURES

## source

1. 18633

/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"

## gene

5921..8765

## CDS

/gene="T24C12.2"

join(5921..6016,6086..6399,6938..7129,7183..7297,  
7349..7544,7595..7794,7842..7970,8020..8270,8322..8427,  
8475..8765)  
/gene="T24C12.2"

/note="Similar to GTPase activating protein (GAP)."

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA82387.1"

/db\_xref="PID:g1086756"

/db\_xref="GI:1086756"

/translation="MVLSCRYVDCGGIKRENOTLLLVSTHNSSEVGLSIDAEK  
RSKEWLEMPGCAEGMLSLKTLMDLKGINSVHGVRVVDENKSGPAKMYLSRK  
ANDEGGEDGGIDGATVVTQIDHILSMOYKKPLDLLFLAGDVQPTASLAVI  
EALPKVELGPSVSLVELMADSRIPIVSLVNSTLKCGDENTLFGOSTSGKMLF  
EILTYGKMLITTLKPYVDKIKERNCYDARAVAGSLKENTRNLLVYOMLE  
KVTSTSTNCPLILKOLYLDNRNVGHSSSGVORLAVSSVFMREFAAILPMPKAE  
IRLDQDPLRVSRITLLSKLRLSNCSVEGSLSEKLELNVFVFEVSEOKSWA  
SFNDNISLVDRSEPOKCTVEKGNLQOVDRSLAMKVAHYKRYQLNTLTIKOK  
DVOCAPRGVYPLSDIKFVNVDKNIITIVETQLEAGVGEATDVAIRORNR  
AAHEIAETGEHFVDAERHVDKIHLLIYVRETMLEPQOLSNVELDKTDELK  
ASYVEERQSHKDSLATICSTIDVTDAIQLAHTEKEKNSRKIDALITTKAKROK  
SESVK"

## gene

complement(10021..17256)

## CDS

/gene="T24C12.3"

complement(join(10021..10249,10474..10732,10788..10902,  
11124..11205,12330..12409,15353..15409,15576..15700,  
15747..15832,15916..16036,16089..16322,16453..16564,  
16609..16707,16982..17057,17102..17256))  
/gene="T24C12.3"

/note="coded for by C. elegans cDNA yk41a1.3; coded for by  
C. elegans cDNA yk128e3.3; coded for by C. elegans cDNA  
yk37c8.3; coded for by C. elegans cDNA CEBSB01F; coded for  
by C. elegans cDNA CEBSG45F; coded for by C. elegans cDNA  
CEBSG45R; coded for by C. elegans cDNA yk37c8.5; coded for  
by C. elegans cDNA yk128e3.5; coded for by C. elegans cDNA

## gene

## CDS

/gene="T24C12.1"

join(12562..12637,12688..12820,12877..13051,13117..13302,  
13855..13936,13984..14053,14107..14262,14312..14570)  
/gene="T24C12.1"

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA82389.1"

/db\_xref="PID:g1086758"

/db\_xref="GI:1086758"

/translation="MNQKACSPGSGVPEASATQSFILFEPSALFLSOSILVEREK  
KQTCARCVFSCSFSTKTFPNSCFMRGKAMLGSMNLLALFLGLGATIVLT  
SPLEVLPIKNDQPVYVGYCFINICLALISIVAIRKQIGVAVVEKICQTLIT  
DAVWVFASINSTASTLSYSHQVHNSKELCSYWDVYQTLICPEPEVQOTCAD  
FLNVTFSFPTGSHNDYCIHLKRLHNSNTEIVCLAFVYTPPLKLMALARDIKV  
ETEMALEYVSEMINEXEDRDGXYIESFNSINMLADSTQTSQTSIRVNSPEYSPFL  
KHSVIEVEVEFVGRDELREEDQALAAHNRQLAAVTEGF"

## gene

## CDS

/gene="T24C12.4"

complement(join(17658..17810,17859..18036,18102..18282,  
18336..18480))  
/gene="T24C12.4"

/codon\_start=1

/evidence=not\_experimental

/protein\_id="AA82390.1"

/db\_xref="PID:g1086759"

/db\_xref="GI:1086759"

/translation="MREYQNRNDHCHILTVRAMFRRLGTFSTIPOLSVMMKYAVPESA  
IPIHIVDAGFQISQKPLKXKENTITSELSLGAAPPCGSDITISACRLQOH  
PSKFLDRCATDADFSLIQCCNTGMSADREYLLIQAGKSNQCFDRGSEFCRFL  
KREDVWGKQSWSCDSTAHAFRICRKTNCFCRDYIRTPLGREFEVSCKGKPVLLIPN"

## BASE COUNT

5810 a 3470 c 3227 g 6126 t

## ORIGIN

## Query Match

81.0%; Score 17; DB 36; Length 18633;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## y

1 tgggataataattgga 17

## Db

12129 TGGGATAAATAATTGA 12113

## RESULT

8

## BSI6086

BSI6086

## LOCUS

BSI6086

## DEFINITION

BSI6086

## ACCESSION

BSI6086

## NID

BSI6086

## VERSION

BSI6086

## KEYWORDS

BSI6086

## SOURCE

BSI6086

## ORGANISM

BSI6086

yk41a1.5; similar to E. coli hypothetical 32.9 kd protein  
in NFO-FRUA intergenic region (SP:YEIN\_ECOLI.P33025)"  
/codon\_start=1  
/protein\_id="AA82388.1"  
/db\_xref="PID:g1086757"  
/db\_xref="GI:1086757"  
/translation="MCSRFQGRSFSKFLQISPOYKEALASGEVVALESTVITHGL  
PYHNLRYTARSLEQRRSSGSHPATIALFDGRIHYGLDDEKLELLASSOMAVYSR  
DIKTLIKKEGGTVAATSMKIHAAIGVFATGGVGRGADOTEDVADQLELSO  
TPCVYVSGVKSITDIPKVEYLFETHSVNCIVGQNVPSFPTSDKACQCESTL  
EEVYHLTKSKSLGEPYGTITLACPIPEKIAADODIQAIDAQVQAIQNTISQSVT  
PFLAVANETLQGASATINIALENNASJAGRLAACORPLTISQSPPTASITKRP  
KVSISATIVDEPAITSDEVDKDGSGYNGVVOVMGVAARNADALARLGCDSVFISA  
IGDNDNGKFFRQNSKRIYESNEDESPDILDSVLSPVAVARVLEIAKHDKQVWL  
EPTDIDKRVKFEPTGLGAVTASPNANEFELKAKLCHVSDPSVITDAGVLELEK  
EKTKLLNTSIFVTLANKGSANVYRNKIGOLEPOSTLPPQLMKNVSVSGADSDNS  
GYTAGLAHNTYVESIQIGQECARLTLQTSLSAIDITRNLLA"

BSI6086 1528 bp DNA BCT 07-MAY-1999  
Bacillus silvestris 16S rRNA gene.  
16S ribosomal RNA; 16S rRNA gene.  
Bacillus silvestris.  
Bacillus silvestris  
Bacteria; Firmicutes; Bacillus/Clostridium group;

REFERENCE 1 (bases 1 to 1528)  
 AUTHORS Rhelms, H., Fruhling, A., Schumann, P., Rohde, M., and Stackebrandt, E.  
 TITLE *Bacillus silvestris* sp. nov., a new member of the genus *Bacillus*  
 that contains lysine in its cell wall  
 JOURNAL Int. J. Syst. Bacteriol. 49 Pt 2, 795-802 (1999)  
 REFERENCE MEDLINE 9253063  
 TITLE 2 (bases 1 to 1528)  
 AUTHORS Rhelms, H.  
 JOURNAL Direct Submission  
 Submitted (14-MAY-1998) Rhelms H., DSM - Deutsche Sammlung von  
 Mikroorganismen und Zellkulturen, Molecular Systematics,  
 Mascheroder Weg 1b, 38124 Braunschweig, FRG  
 Revised by author 16-NOV-98

REMARK FEATURES  
 source Location/Qualifiers  
 1..1528  
 /organism="Bacillus silvestris"  
 /isolate="HR3-23"  
 /db\_xref="taxon:76853"  
 22..1528  
 /gene="16S rRNA"  
 /product="16S ribosomal RNA"  
 22..1528  
 /gene="16S rRNA"  
 BASE COUNT 391 a 351 c 463 g 323 t  
 ORIGIN

Query Match 80.0%; Score 16.8; DB 1; Length 1528;  
 Best Local Similarity 90.0%; Pred. No. 2,1e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgggataataatcgaagc 20  
 |||||||||||||  
 Db 1466 TGGGATAATAATATGCGGTC 1485

RESULT 9  
 T1F15 55721 bp DNA PLN 10-SEP-1998  
 LOCUS Arabidopsis thaliana chromosome 1 BAC T1F15 sequence, complete  
 DEFINITION sequence.  
 AC004393  
 NID 93108024  
 VERSION AC004393.1 GI:3108024  
 KEYWORDS HTG.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euhylophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsis.  
 1 (bases 1 to 55721)  
 Vysotskaya, V.S., Schwartz, J.R., Toriumi, M., Yu, G., O'H., O.,  
 Kwan, A., Liu, S., Li, J., Araujo, R., Au, M., Brendel, V., Buehler, E.,  
 Conway, A.B., Conway, A.R., Dewar, K., Feng, J., Kim, C., Kutz, D.,  
 Li, Y., Palm, C.J., Shinn, P., Sun, H., Davis, R.W., Ecker, J.R.,  
 Federiguel, N.A. and Theologis, A.  
 Arabidopsis thaliana chromosome 1 BAC T1F15 sequence  
 Unpublished (1998)  
 This sequence is of BAC T1F15 from Arabidopsis thaliana chromosome  
 1. The sequence does not represent the sequence of the entire  
 insert of this clone. It is shorter by 11342 bp because we submit  
 only the unique sequence of the clone. However, in order to  
 facilitate the joining of overlapping clones in the future for  
 creation of larger contigs, we provide a small overlap (200 bp)  
 between overlapping submitted clones. The 3' end of this sequence  
 overlaps by 200 bp the 5' end of the sequence of the BAC FIN21.  
 2 (bases 1 to 55721)  
 Theologis, A.  
 Direct Submission  
 Submitted (11-MAR-1998) Plant Gene Expression Center, 800 Buchanan  
 Street, Albany, CA 94710, USA

REFERENCE 3 (bases 1 to 55721)  
 AUTHORS Theologis, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-MAY-1998) Plant Gene Expression Center, 800 Buchanan  
 Street, Albany, CA 94710, USA  
 REFERENCE 4 (bases 1 to 55721)  
 AUTHORS Theologis, A.  
 JOURNAL Direct Submission  
 Submitted (03-JUN-1998) Plant Gene Expression Center, 800 Buchanan  
 St., Albany, CA 94710, USA  
 On May 2, 1998 this sequence version replaced gi:2951938.

COMMENT FEATURES  
 source Location/Qualifiers  
 1..55721  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="1"  
 /clone="T1F15"  
 1531..4576  
 /gene="T1F15.1"  
 join(1531..2665,2775..3044,3176..3351,3423..3633,  
 3708..3945,4015..4168,4259..4576)  
 /gene="T1F15.1"  
 /note="Strong similarity to receptor kinase gblM80238 from  
 A. thaliana."  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="AAC18783.1"  
 /db\_xref="PID:g317659"  
 /db\_xref="GI:317659"  
 /translation="MCSNGIFVSLTLTLGLKSCSETDTLHOGFLLKDGQELVSAFK  
 IFLKFEFNKSENLVGIWENMLYLTDSQDPVWIANRNPISDRSGSLVDSGLR  
 LKILRGASTMLELSTETRTNTLLDSCNLOLEMDADGSKRVYMOSEFDTDL  
 LPKCKLGFDPDKTRKRWELTWLGDPLDPASSYPFGMDPTNTNTNLTLMRNMWSSGL  
 WNGRSEELNECGLEFSEVSTKSGYETWSDQDDARTFFPLIMDEGILLRSDM  
 HRONRONRYNRNCLAGVYVRDEPYGTFSTRVYSSASNGSVLSCGFSSVSAIC  
 LONSCLAVASTEPDGTGCEIWNVTYPRNKGSASHSPRTIYIRNGQENKVAAMHIV  
 ALLEMLPIWFIYLVLRKFNWGRNCIITKRTVLSWVFLTSSPFLPIODV  
 FYFVEYTFYFGESGLKHQEMLELRIGIDRSCIHKNRKNSENNLOIFSESVSAT  
 DDPSENKLEGGGPPYKRLKNGEVAIAKRSLASGQGLVFKNALIIIAKQHTN  
 LVQVLGCCIEKDEKMLIETWQNKSLDTFLFDPURKVLWMTLRFRTIMETIGGLTL  
 HKYSRLVYHRIKASNIILDEDMNPKISDFGLARLFGAETRAKTRVAGTGYNSP  
 EYFREGLEFSKSDVFSFVGLMLEITICGRKNNSFHHLEGPLIVHWNLFKENKIRE  
 VIDLSRSDALDYPOVLCVQVALLCQENAEPRSMLDVSMIYGBGNALSLPKRP  
 AFYDGRSRSPENKVEPQEPENVSASITITVLEAR"  
 7204..9911  
 /gene="T1F15.2"  
 join(7204..8692,9241..9911)  
 /gene="T1F15.2"  
 /note="Similar to ERECTA receptor protein kinase gblU47029  
 from A. thaliana."  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="AAC18784.1"  
 /db\_xref="PID:g317660"  
 /db\_xref="GI:317660"  
 /translation="MTLSLLHLHLHLTLMTSPSLSPDGIALLSIKSAVNDSSSA  
 FSDWMDNTPPCRHSGISGCMNIDSSSTRVYGSILAKKHLRGITPESGLIYLRRLN  
 LHNNEELGSLPTOLFNTSLHSIFLYNNLSGTLPSICLPKQLNDLSMSLSGL  
 SPDLNCKOGLRLSLNNNSGELIPGDIWELNLADLSANFSEIDIGELKS  
 LSGTLNLSFNLHSGQIPNSLNPVYVLDLRNDFSGEIPQSGSFQGPFLANPL  
 KICGFLPKTKCDKDENSPGRKSPENNASRSLGSLGLVLSVADAAVAFFGLV  
 VLYLWKKRDEGGSCGCMKIGGSYKSCCITCGPREDSAEAGNRRGGKGG  
 ELVAIDGFEDELIRASAVYAGKGLTIYKVVYAGNCPVAVRRLGSGEGRYKE  
 FVEVQMGVYKHNVVKLRAVYVADPEKLLIDFVNGSLAALRNRNQSPSLTW  
 STRIKIRKAGARGLAYLHESPRLVHGVKPSKILDSSTFTYISDFGLRIITTA  
 ASASNNPSSSSAAGELGALPYTSIKPDRSGYKAPARLPGCPGTQKWPYVSFG  
 VVLMELTLGKSPSSSTSVYVYVVDVYKWKVGGFEETPLSDMDVPMILQYV  
 HAKQOVLSVPHLALACREBDPVRPRKKNVSENIDK"  
 10674..19681  
 /gene="T1F15.3"  
 join(10674..10785,10946..11028,11116..11173,11269..11318,

11488..11607,11671..11721,11831..11899,12316..12459,  
12547..12730,12809..12900,12979..14599,14801..14992,  
15191..15543,15849..15991,16076..16658,16735..16881,  
17063..17335,17501..17608,17692..17897,18107..18410,  
18496..19113,19508..19681)  
/gene="TIF15.3"  
/note="Similar to putative DNA polymerase gblM29683 from  
S. cerevisiae."  
/evidence=not\_experimental  
/protein\_id="AAC18785.1"  
/db\_xref="PID:g3176661"  
/db\_xref="GI:3176661"  
/translation="MADSGSNVFSLRIVSIDYVNASPIPGNYICYSFQSEVNEV  
PVIRIGSTPAGCTCCLHIALPYLIPCEIPLEHKKGDSSTLASELKALV  
KGNASRHOIHDECEIYRAKFEYGHSTEAFFKITYKPSYHPDPVAAASLLAG  
AVGKSLRPIHESIPILOFLVDYNYLGMGHVHSMKRPVPHRRPRDLODCR  
GORIDEPATIRANSSAAVSPYWSLSTIPGOMNMLSESDTPLSQSHQHNR  
RQSLCELEGATSDILNQFKMNSIQSDTMVQSLVAILWEERYERTGVADAI  
PPDGRKSADVLTQMSDYVFGNMLKEMNLVELSPGKMPVASAGDMAKPEI  
TDLQALNHMGTCSEFPASQLSPGKSEASAMENDEYKPTDRTDPAIODAEI  
GLFKFASQAADINSDELIRETILSPILPLASIKVLEMASTVDVSGQSCODI  
LDSOENLPDGSSTKRALPSNPSONRTSDQSLEIYASVDPDSTNGASNSF  
RRYKSDLHISEWETKNSFSKSNKSNVWGGLPTLTKNLKQDSTNADSKGL  
TKISSIPNEMTONTYVYKHOADVNTIDRNLACSLADLRRKRLKRGSPSQ  
HMSKRYDRSRHNEKCTLRCEAKKQGPALSAESEFVCGDTPNLSIDGCECNI  
STESSELHSDRCASKEATASNDENVLRNLSSTVPKGPQVYESLTASNHNGI  
EIDSVQSGREOSTAENETDEGTGLITLSKRPDILDCSAGIDASHHETARPK  
QHDYEGNSNDIPFPLEDNKKEKHPQTSGLPIPHANDSNLITLTPASPSY  
DSVQWISNDKGSNIDSEKOPLDNNDHNDGASTDLASNVASVEHEQNNHNV  
NSSMAYTESEIDLPKGTFLNLTNLQASVQELSQISPGKSGPPLSQMGFRPAS  
MGAGQOULTLSIEHAESRGDLRDPDSDENVYALVQNDSEFAVEFLVLESPSI  
DORNDLQSCCKLVLEEROLFREFYETLCMDPVLALMDLQSGSGLIARAAOL  
GIRFLNNSRTSPPTTNSNDNRKIGNTLDPDLYANOVVEVEDWEDWTHSG  
VHYGGRVLANPRLIGEVLNMTITAVSEAVLRQVPSIPKVLTERSSGPACAR  
YCIETVIRANLNLLEMSOLDMNRSEIARFGIDFVSLSQSYRESMLRLA  
HTONYLAIISQNOVASOPAMECPVLMPESEAYDIPVTLDFOSQSYRESMLRLA  
ESTLCGLAHLKNNLTGVSSYSLDLVQDNLQTLQPNVMVYPPVRRGILPRLLE  
EILSTRIMVKKAKKLTPSEAVLHRIINAPOLAKLTANTYGYTAVFGSRGICATL  
ADSLVOCGRSTLEKATISFVANNDNMNARVYGGSDGRTVKEATVAVGOETAS  
AITEMNHPYTLKMEKYHPCFLTKRKYGYSTESNOQEPFIDAKGICTVRYDCE  
AVAKMEQSLRLEFEONKISKVSSTLYRKRRLISGVSQDITFAEVNLYGTSTD  
SSLPAPAIATKSMKADPTEPRYARVYVYHGGPRLALVMDVDPVLVDVDP  
YRLNDLYINKOITIPALORVFGVGLNVTFLSEMPORLSEKNSHKTETI  
DYFLSKHICLGEVQESALQCNRCNCSAAATVMTKSEKRMOLATCRIC  
GGDWMVQSGCKNSLACSVFYERKQKRLRGSLATSELYPKMAWPF"  
20807..26755  
/gene="TIF15.4"  
/join(20807..21106,21194..21267,21348..21599,21949..22159,  
22261..22350,22448..22544,22957..23039,23401..23510,  
23630..23751,23901..24008,24085..24158,24510..24635,  
24866..24953,25260..25435,25567..25711,25831..25980,  
26086..26177,26254..26363,26473..26536,26611..26755)  
/gene="TIF15.4"  
/note="Similar to mannosyl-oligosaccharide glucosylase  
gblX87237 from Homo sapiens."  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAC18786.1"  
/db\_xref="PID:g3176662"  
/db\_xref="GI:3176662"  
/translation="MTGASRRSAGRIKSSLSLSPGSDGSAVPPSIRNGKELVSG  
AFKTNLILGLIIGIIVYFNLRVRLGLLEDSQGRVTPFPAPVRLSMQ  
GEHRESLYMGTYRPHYFEGVARPLSLVAGMLGVCXKDMYVRRHCNSDLSTG  
WRHNGRDYRQELVENDVETISFVSKGDLCTGDMVARDVKKKGFVAVASAS  
GSAAGGLTFVWRSSADAHDTILLVLDLQNLQLEHVGKFTPHYNSLDVQNL  
ALQARKFRLQSLDSTEDSSNIYIFVYSLNKLDSGLPFTIDIPFISIGKSSNE  
KRLTSLGLDLKHHOEFDAKFNCEKLSKHDSETLGVRTAIANMLGSSINE  
YCGSKIYVPKSTQSRDNFLVWPALEYAVPSPRPGFLGMDDEDFLLIWRMI  
RITLDIVGHMLDLINDGWI PROIIGALSLVPEFVVOYPSNGSPPLFLIWRMI  
IDAIEMKEFVASEKDEVLSTFLERASVLDNMFQFNTSQKKEIGSIFMGRDVTITQ  
ELNPKTUSGLDDTIPRASHPESEDERHVDLCWYTLADCMHSITTELLGKEDKLSVST  
RMKRCQGFQSHFNIFLFLITLITLISHSWIISBOMHYDSDGYAFDFGHNTEKXLI

gene  
CDS  
FEATURES  
source  
BASE COUNT  
ORIGIN  
Query Match  
Best Local Similarity  
Matches  
Query  
Db  
RESULT  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
BASE COUNT  
ORIGIN  
WHEVIOENQSLQVLRKTEGKPKLVLPHLGIVSEFPMSRIIPDPSILEKQDLI  
SNSNISIMSYDGLSLAKTSMYKRNTEHDAPWRPIMNNMYMLISLYHYSDG  
PYREKRAKATYELSTRSLNIRNVNVEYETGYMEQYDQVKGKGTFLTGMWMLTLLI  
MSEDYVIF"  
complement(28859..30381)  
/gene="TIF15.5"  
complement(28859..30381)  
/gene="TIF15.5"  
/note="Contains similarity to beta scrin gbl247541 from  
Limulus polyphemus. ESTs gblIT04493 and gblAA58595 come  
from this gene."  
/evidence=not\_experimental  
/protein\_id="AAC18788.1"  
/db\_xref="PID:g3176664"  
/db\_xref="GI:3176664"  
/translation="KCMRRNNITLTVIDSNEALIGMEHEDLSARHOYSVLTAPFS  
ATYLSVQKIGLWQAFALSGKRRIVNHGCEKGNLDLSRSENFMDDPLTIGLPD  
DVAKQCLAVPRARFPMGSGVCKRMPFVQSKFEIVRRLAGMLEMLVYLTNNAGK  
DNREWMDICIGKSLSPMPGPAKGFVYVVDGKLVYIACGCMINGSLVASADYQ

Query Match 80.0%; Score 16.8; DB 34; Length 144514;  
 Best Local Similarity 90.0%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gggataataatggagcgt 21  
 ||||||| |||||||  
 Db 71688 GGGATTAATAATGGAGCCT 71669

RESULT 11  
 AC007437/c AC007437 211912 bp DNA HTG 17-JUN-1999  
 LOCUS Homo sapiens, WORKING DRAFT SEQUENCE, 31 unordered pieces.  
 DEFINITION AC007437  
 ACCESSION AC007437  
 NID 95091576  
 VERSION AC007437.15 GI:5091576  
 KEYWORDS HTG, HTGS, PHASE1.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 211912)  
 Muzny, D., Arenson, A.D., Brundage, E., Carvelli, K., Chen, E., Di, W.,  
 Ding, Y., Dugan, S., Durbin, J., Forcum, J., Ganesh, R., Garcia, C.,  
 Goodman, M., Gorrell, J.H., Haywood, W., Jackson, L., Kampal, R.,  
 Karpathy, S., Leal, B., Li, Y., Liu, W., Logan, O., Lu, J., Ly, T.,  
 Martinez, C., Oswal, G., Perez, L., Rashid, N.D., Rowland, K.,  
 Savage, L.C., Scherer, S., Shen, H., Timms, K.M., Todd, J., Vo, Q.,  
 Worley, K.C., Yu, W., Chinault, C., Nelson, D. and Gibbs, R.A.

TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 211912)  
 AUTHOR Worley, K.C.

COMMENT Submitted (30-APR-1999) Molecular and Human Genetics, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jun 17, 1999 this sequence version replaced gi:5080739.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 31 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 83103: contig of 83103 bp in length  
 \* 83104 83123: gap of unknown length  
 \* 83124 163418: contig of 80295 bp in length  
 \* 163419 163438: gap of unknown length  
 \* 163439 187893: contig of 24455 bp in length  
 \* 187894 187913: gap of unknown length  
 \* 187914 189387: contig of 1474 bp in length  
 \* 189388 189407: gap of unknown length  
 \* 189408 190763: contig of 1356 bp in length  
 \* 190764 190783: gap of unknown length  
 \* 190784 192044: contig of 1261 bp in length  
 \* 192045 192064: gap of unknown length  
 \* 192065 193272: contig of 1208 bp in length  
 \* 193273 193292: gap of unknown length  
 \* 193293 194416: contig of 1124 bp in length  
 \* 194417 194436: gap of unknown length  
 \* 194437 195486: contig of 1050 bp in length  
 \* 195487 195506: gap of unknown length  
 \* 195507 196528: contig of 1022 bp in length  
 \* 196529 196548: gap of unknown length  
 \* 196549 197562: contig of 1014 bp in length  
 \* 197563 197582: gap of unknown length  
 \* 197583 198556: contig of 974 bp in length  
 \* 198557 198576: gap of unknown length  
 \* 198577 199515: contig of 939 bp in length  
 \* 199516 199535: gap of unknown length  
 \* 199536 200421: contig of 886 bp in length  
 \* 200422 200441: gap of unknown length

200442 201325: contig of 884 bp in length  
 \* 201326 201345: gap of unknown length  
 \* 201346 202224: contig of 879 bp in length  
 \* 202225 202244: gap of unknown length  
 \* 202245 203105: contig of 861 bp in length  
 \* 203106 203125: gap of unknown length  
 \* 203126 203869: contig of 744 bp in length  
 \* 203870 203889: gap of unknown length  
 \* 203890 204628: contig of 739 bp in length  
 \* 204629 204648: gap of unknown length  
 \* 204649 205377: contig of 729 bp in length  
 \* 205378 205397: gap of unknown length  
 \* 205398 206080: contig of 663 bp in length  
 \* 206081 206100: gap of unknown length  
 \* 206101 206782: contig of 682 bp in length  
 \* 206783 206803: gap of unknown length  
 \* 206803 207472: contig of 670 bp in length  
 \* 207473 207492: gap of unknown length  
 \* 207493 208158: contig of 666 bp in length  
 \* 208159 208178: gap of unknown length  
 \* 208179 208830: contig of 652 bp in length  
 \* 208831 208850: gap of unknown length  
 \* 208851 209485: contig of 635 bp in length  
 \* 209486 209505: gap of unknown length  
 \* 209506 210104: contig of 599 bp in length  
 \* 210105 210124: gap of unknown length  
 \* 210125 210714: contig of 590 bp in length  
 \* 210715 210734: gap of unknown length  
 \* 210735 211261: contig of 527 bp in length  
 \* 211262 211281: gap of unknown length  
 \* 211282 211706: contig of 425 bp in length  
 \* 211707 211726: gap of unknown length  
 \* 211727 211912: contig of 186 bp in length.

FEATURES  
 source location/Qualifiers  
 1..211912  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 63141 a 41899 c 41859 g 63493 t 1520 others  
 ORIGIN

Query Match 80.0%; Score 16.8; DB 35; Length 211912;  
 Best Local Similarity 90.0%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gggataataatggagcgt 21  
 ||||||| |||||||  
 Db 11517 GGGATTAATAATGGAGCCT 11498

RESULT 12  
 AC007656/c AC007656 171226 bp DNA PRI 29-MAY-1999  
 LOCUS Homo sapiens 12q22 BAC RPEC11-534P6 (Rowswell Park Cancer Institute  
 DEFINITION Human BAC Library) complete sequence.  
 ACCESSION AC007656  
 NID 94926854  
 VERSION AC007656.2 GI:4926854  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 171226)  
 Muzny, D., Arenson, A.D., Bala, S., Bouck, J., Bunac, C., Chen, Z.,  
 Ding, Y., Dugan, S., Durbin, J., Dutta, S., Forcum, J., Garcia, C.,  
 Gorrell, J.H., Gorrell, L.L., Hernandez, J., Issar, A., Jackson, L.,  
 Kneitz, S., Kondewski, N., Lau, S., Leal, B., Lee, E., Lichtarge, O.,  
 Liu, W., Logan, O., Lu, J., Maronde, I., Martinez, C., Merscher, S.,  
 Miller, A., Montgomery, K., Murty, V.V., Oswal, G., Pampell, L.R.,  
 Parish, B.J., Perez, J., Rashid, N.D., Renauld, B., Rives, C.,  
 Scherer, S.E., Shen, H., Shim, M., Simon, M., Vo, Q., Williamson, A.,  
 Worley, K.C., Yang, R., Yu, W., Zhou, X., Chagnant, R.S.K.,

TITLE Kucherlapati, R., Nelson, D. and Gibbs, R.A.  
 JOURNAL Direct Submission  
 REFERENCE 2 (bases 1 to 171236)  
 AUTHORS Morley, R.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-MAY-1999) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 171236)  
 AUTHORS Morley, R.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-MAY-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On May 29, 1999 this sequence version replaced g1:4895152.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

## COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT.

FEATURES  
 Source Location/Qualifiers

1. 171236  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RPC111-534P6"  
 /chromosome="12q22"  
 repeat\_region complement(3. .266)  
 /rpt\_family="MLT1J"  
 repeat\_region complement(562. .663)  
 /rpt\_family="L2"  
 repeat\_region complement(807. .979)  
 /rpt\_family="L2"  
 repeat\_region complement(1094. .1472)  
 /rpt\_family="MLT1A2"  
 repeat\_region 1552. .1644  
 /rpt\_family="MIR"  
 repeat\_region complement(1705. .1915)  
 /rpt\_family="L2"  
 repeat\_region complement(2410. .2495)  
 /rpt\_family="MIR"

repeat\_region 2895. .3253  
 /rpt\_family="L2"  
 repeat\_region 3369. .3405  
 /rpt\_family="(TAA)n"  
 repeat\_region complement(4021. .4193)  
 /rpt\_family="MER117"  
 misc\_feature 4280. .4704  
 /note="Region: zh63e12.s1 Homo sapiens cDNA W86760"  
 STS 4556. .4680  
 /standard\_name="G23213"  
 /db\_xref="dbSTS:37051"  
 repeat\_region 5227. .5303  
 /rpt\_family="L2"  
 repeat\_region 5925. .6078  
 /rpt\_family="MIR"  
 repeat\_region complement(6510. .6694)  
 /rpt\_family="MIR"  
 repeat\_region 8159. .8179  
 /rpt\_family="AT\_rich"  
 repeat\_region 9049. .9090  
 /rpt\_family="(CA)n"  
 misc\_feature 9089. .9090  
 /note="From two clones. The rest of the clones missing AC."  
 /function="Ambiguities"  
 repeat\_region 9103. .9174  
 /rpt\_family="MER3"  
 repeat\_region 9175. .9508  
 /rpt\_family="MER1B"  
 repeat\_region 9509. .9643  
 /rpt\_family="MER3"  
 repeat\_region 9646. .9736  
 /rpt\_family="MIR"  
 repeat\_region complement(10021. .10285)  
 /rpt\_family="MLT1K"  
 repeat\_region 10671. .10915  
 /rpt\_family="MLT1J1"  
 repeat\_region 11573. .11600  
 /rpt\_family="AT\_rich"  
 repeat\_region 12117. .12260  
 /rpt\_family="L1ME"  
 repeat\_region 12361. .12905  
 /rpt\_family="L1ME"  
 repeat\_region 12906. .13187  
 /rpt\_family="AluSx"  
 repeat\_region 13188. .13238  
 /rpt\_family="(CCA)n"  
 repeat\_region 13241. .14273  
 /rpt\_family="L1ME"  
 repeat\_region 14326. .14548  
 /rpt\_family="L1M4"  
 repeat\_region complement(14549. .14593)  
 /rpt\_family="MER5B"  
 repeat\_region complement(15349. .15476)  
 /rpt\_family="MER5A"  
 repeat\_region 15851. .16035  
 /rpt\_family="MIR"  
 repeat\_region 16496. .16781  
 /rpt\_family="AluSx"  
 repeat\_region 16782. .16842  
 /rpt\_family="(TAAA)n"  
 repeat\_region 19115. .19173  
 /rpt\_family="polyprymidine"  
 repeat\_region 19199. .19228  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(20685. .20793)  
 /rpt\_family="MER5B"  
 repeat\_region 23191. .23329  
 /rpt\_family="L2"  
 repeat\_region 23534. .23566  
 /rpt\_family="MIR"  
 repeat\_region 23574. .23768  
 /rpt\_family="L2"

```

repeat_region      23839..23957
                    /rpt_family="L2"
repeat_region      24731..24816
                    /rpt_family="L2"
repeat_region      25576..26264
                    /rpt_family="L2"
repeat_region      26301..26332
                    /rpt_family="L2"
repeat_region      26333..26562
                    /rpt_family="MER46A"
repeat_region      26563..26655
                    /rpt_family="L2"
repeat_region      26755..27100
                    /rpt_family="L2"
repeat_region      27454..27590
                    /rpt_family="L2"
repeat_region      27591..28133
                    /rpt_family="MER34B"
repeat_region      28144..28252
                    /rpt_family="MIR"
repeat_region      28597..28933
                    /rpt_family="L2"
repeat_region      28934..29214
                    /rpt_family="ALUSg"
repeat_region      29215..29371
                    /rpt_family="L2"
repeat_region      30072..30213
                    /rpt_family="L2"
repeat_region      30273..30365
                    /rpt_family="ALU/FRAM"
repeat_region      30657..30779
                    /rpt_family="MLT1A1"
repeat_region      32217..32245

```

Query Match 80.0% Score 16.8: DB 42; Length 171236;

Best Local Similarity 90.0% Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 gggaataatctgagcgt 21

Db 151435 GCGAATAATTATGTGAGCCT 151416

```

RESULT 13
AC004806/c 129440 bp DNA PRI 20-JAN-1999
LOCUS Homo sapiens 129424.2 BAC RPC111-360511 (Roswell Park Cancer
DEFINITION Institute Human BAC Library) complete sequence.
ACCESSION AC004806
NID g3810570
VERSION AC004806.1 GI:3810570
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 129440)
AUTHORS Munzly,D., Arenson,A.D., Bouck,J., Bunac,C., Chang,J., Chen,Z.,
Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,
Gorrell,L.L., Hernandez,J., Jackson,L., Kondejewski,N., Lau,S.,
Leal,B., Lee,E., Lichtarge,O., Liu,W., Logan,O., Lu,J.,
Marandel,L.R., Martinez,C., Merscher,S., Montgomery,K., Oswal,G.,
Pampell,L.R., Parish,B.J., Perez,L., Rashid,N.D., Renault,B.,
Rives,C., Scherer,S.E., Shen,H., Simon,M., Vo,O., Williamson,A.,
Worley,K.C., Yu,W., Zhou,X., Kucheriapatti,R., Nelson,D. and
Gibbs,R.A.
Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 129440)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

```

```

REFERENCE 3 (bases 1 to 129440)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 129440)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 30, 1998 this sequence version replaced gi:3327896.
INFORMATION: http://gc.bcm.tmc.edu:8088/home.html or email
gc-help@bcm.tmc.edu

```

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:  
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of low estimated quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT-----

```

----- Summary Statistics -----
Contig length: 172045
Phrap values in estimate: 83363
Average error rate (BCM-Phrap estimate): 2.00599e-06
Fraction of Phrap values less than 40 : 0.00133153
Number of consensus changing edits: 5
Number of N's in consensus : 0

```

```

----- Consensus changing edits -----
Position Original/Context Edited/Context
9274 ttttccctt(n)ccctctctc ttttccctt(c)ccctctctc
15955 cttttttt(n)tcctcgaag cttttttt(t)tcctcgaag
51005 aaaaatttga(n)ataatttgg aaaaatttga(a)ataatttgg
51240 acggtgaac(n)ctgtctctac acggtgaac(c)ctgtctctac
168300 gctctctct(n)accgcccaaga gctctctct(c)accgcccaaga

```

```

----- Low Quality Bases -----
Quality Position Surrounding Sequence

```



```

11      161149      gtggagggtg(t)gaggaaggag

----- Distribution of Quality < 40 Bases -----

# bases
1001
901
801
701
601
501
401
301
201
101
01
-----
          *           *           *           *           *
5         10        15        20        25        30        35        40
Phrap Value Range

FEATURES
Version: 1.0 pfto.
Source Location/Qualifiers
1. .129440
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RPC111-360E11"
/chromosome="12q24.2"
/complement(1173, .1511)
/rpt_family="L2"
complement(1593, .1629)
/rpt_family="L2"
1634, .1702
/rpt_family="L2"
1874, .1962
/rpt_family="L2"
complement(2217, .2339)
/rpt_family="L2"
2819, .3059
/standard_name="D12S2452"
/db_xref="GDB:9785966"
complement(5117, .5446)
/rpt_family="AluYb8"
5802, .5970
/rpt_family="MER3"
7454, .7654
/function="Low coverage"
8488, .8815
/rpt_family="L2"
10924, .11031
/rpt_family="L2"
complement(11098, .11571)
/rpt_family="MT1P"
12021, .12190
/rpt_family="MIR"
12376, .12614
/rpt_family="MIR"
13228, .13620
/rpt_family="MT1A1"
complement(14782, .14900)
/rpt_family="(TGA)n"
14874, .15352
/function="Low coverage"
complement(15012, .15094)
/rpt_family="(TGA)n"
complement(15096, .15215)
/rpt_family="(TGA)n"
15353, .15402
/note="ct rich region"
/function="Gap in sequence of less than 50 bases
containing T(1C)(t) repeats."
15403, .156172
```

```

repeat_region      /function="low coverage"
                   complement(16381..16500)
                   /rpt_family="L2"
repeat_region      complement(16501..16676)
                   /rpt_family="TRIGGER2"
repeat_region      complement(16677..16814)
                   /rpt_family="MER1A"
repeat_region      complement(16847..17260)
                   /rpt_family="Aluub"
repeat_region      17811..17909
                   /rpt_family="Aluuo/FRAM"
repeat_region      complement(18356..18773)
                   /rpt_family="MLT1C"
repeat_region      complement(19048..19352)
                   /rpt_family="Aluuo"
repeat_region      complement(20383..20678)
                   /rpt_family="Alusx"
repeat_region      22039..22383
                   /rpt_family="Aluuo"
repeat_region      complement(22657..22860)
                   /rpt_family="MER3"
repeat_region      24376..24506
                   /rpt_family="MIR"
repeat_region      24591..24730
                   /rpt_family="MLT1C"
repeat_region      24928..25085

Query Match      78.1%: Score 16.4; DB 11; Length 129440;
Best Local Similarity 94.4%: Freq. No. 2e+02; 1; Indels 0; Gaps 0
Matches 17; Conservative 0; Mismatches

QY      4 gataataattgagcgt 21
|||||
Db 12491 GATAATAATTGAGCCTT 12474

```

RESULT	14
AC005547/c	
LOCUS	
DEFINITION	Homo sapiens chromosome 19, cosmid R29228, complete sequence.
ACCESSION	AC005547
NID	93478634
VERSION	AC005547.1
KEYWORDS	GI:3478634
SOURCE	HTG.
ORGANISM	human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 46127) Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Adamson,A.W., Burkhardt-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stillwagen,S., Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Gaines,D., Dangnan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J., Lu,S., Attk,C., Andreise,T., Trankheim,M., Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G., Krommiller,B., Arellano,A., Ov,d., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.: Sequence analysis of an -600 kb human contig in 19q12 between DIS5430 and DIS576 Unpublished 2 (bases 1 to 46127) Lamerdin,J.E.
JOURNAL	Direct Submission
REFERENCE	Submitted (27-AUG-1998) Joint Genome Institute, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
TITLE	Map and sequence oriented from q centromere to telomere. Cosmid R29228 represents the telomeric end of the current sequencing tiling path. R29228 is separated from cosmid R19750 to the right by <1 kb. Additional chromosome 19 map and sequence information may be obtained at: <a href="http://www.bto.llnl.gov/dbfp/genome/genome.html">http://www.bto.llnl.gov/dbfp/genome/genome.html</a> .
COMMENT	Location/Qualifiers  1..46127  /organism="Homo sapiens"
FEATURES	
SOURCE	

```

/db_xref="taxon:9606"
/clone="R29228"
/chromosome="19"
/map="19q12 between D19S430 and D19S776"
/cell_line="5HL2-B"
/clone_11b="L119NC03 R chromosome 19-specific cosmid library"
/note="Cosmid library constructed at LNL from flow-sorted chromosomes from human-hamster hybrid 5HL2-B, which carries chromosome 19 as its only human chromosome."
792. 1092
/rpt_family="AluXs"
1293. 1444
/note="predicted exon, program: grai12exons_human_1.3, frame: 1, quality: good, score: 51.000"
2718. 2839
/rpt_family="MIR"
2842. 3144
/rpt_family="AluSg"
3354. 3546
/rpt_family="AluJo"
complement(3751..4055)
/rpt_family="AluJo"
complement(4188..4276)
/note="predicted exon, program: grai12exons_human_1.3, frame: 0, quality: excellent, score: 78.000"
4525. 5000
/rpt_family="MER1A"
complement(5595..5894)
5924. 6224
/rpt_family="AluJo"
5924. 6224
/rpt_family="AluSx"
8333. 8510
/rpt_family="L1MD3"
9477. 9776
/rpt_family="AluJo"
complement(10802..10850)
/rpt_family="MLT2FB"
complement(10851..10958)
/rpt_family="MER7B"
10982. 11329
/rpt_family="THER1B"
complement(11361..11595)
/rpt_family="MER7A"
complement(11596..11950)
/rpt_family="MLT2FB"
11997. 12060
/rpt_family="LRR16A"
13567. 13605
/rpt_family="(GA)n"
13964. 13990
/rpt_family="(GAA)n"
13992. 14127
/rpt_family="(GAA)n"
14140. 14284
/rpt_family="(GAA)n"
14321. 14607
/rpt_family="AluJo"
complement(14886..15186)
/rpt_family="AluSx"
15207. 15296
/rpt_family="MER81"
complement(15373..15733)
/rpt_family="THER1C"
16557. 16656
/note="predicted exon, program: grai12exons_human_1.3, frame: 0, quality: excellent, score: 90.000"
17248. 17354
/rpt_family="MIR"
/note="predicted exon, program: grai12exons_human_1.3, frame: 0, quality: excellent, score: 79.000"
complement(18564..18597)

```

```

repeat_region
complement(18599..19401)
/rpt_family="L1PAL5"
19487. 19728
/note="predicted exon, program: grai12exons_human_1.3, frame: 0, quality: good, score: 56.000"
complement(20628..20673)
/rpt_family="(CA)n"
complement(20755..20860)
/rpt_family="MIR"
21594. 21817
/rpt_family="LINE2"
22718. 22800
/rpt_family="MIR"
22963. 23049
/rpt_family="LINE2"
23354. 23401
/rpt_family="LINE2"
23703. 24002
/rpt_family="AluSx"
24445. 24543
/rpt_family="(GAA)n"
complement(26091..26146)
/note="predicted exon, program: grai12exons_human_1.3, frame: 1, quality: excellent, score: 80.000"
26883. 27184
/rpt_family="AluSx"
29380. 29424
/note="predicted exon, program: grai12exons_human_1.3, frame: 0, quality: good, score: 57.000"
29608. 29639
/rpt_family="(CA)n"
29854. 29989
/rpt_family="LINE2"
complement(30969..31005)
/rpt_family="AT-rich"
31959. 32049
/rpt_family="(GA)n"
32361. 32670
/note="DSD similarity to overlapping ESTs:-(32361..32670) AA824408 at175607.s1 Soares testis NHT Homo sapiens cDNA clone 1376676 3'; (463..154); 99% identity.--AA971001 op67h02.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1581939 3'; (154..4); 100% identity."
32496. 32670
/note="predicted exon, program: grai12exons_human_1.3, frame: 2, quality: good, score: 71.000"
complement(34259..34296)
/rpt_family="POLY-A"
34300. 34450
/note="DSD similarity to overlapping ESTs:-AA824408 at175607.s1 Soares testis NHT Homo sapiens cDNA clone 1376676 3'; (153..1); 99% identity.--AA971001 op67h02.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1581939 3'; (154..4); 100% identity."
35293. 35377
/note="DSD similarity to AA609196 af12d10.s1 Soares testis NHT Homo sapiens cDNA clone 1031443 3'; (439..356); 99% identity."
35846. 36022
/note="predicted exon, program: grai12exons_human_1.3, frame: 2, quality: good, score: 73.000"
complement(37184..37244)
/rpt_family="(GA)n"
complement(37245..37307)
/rpt_family="AT-rich"
38183. 38246
/note="predicted exon, program: grai12exons_human_1.3, frame: 2, quality: excellent, score: 96.000"
38323. 38424
/note="predicted exon, program: grai12exons_human_1.3, frame: 1, quality: excellent, score: 82.000"

```



/product="p33 protein"  
/protein\_id="CA77163.1"  
/db\_xref="PID:e1375376"  
/db\_xref="PID:94239717"  
/db\_xref="GI:4239717"  
/translation="MEAFASNODLLEKIRRRRTYHRRKFPDDVVRDPTIDVGYD  
RDPYVLADYFSLFFFLNNDVNGPLAAIVSPYNGYKTRAHNDNPNEGATYI  
KTDKNRAFLIKAPAPMRKGYISFRATLADVSESEFICSPVGSFKCCQTIS  
ENLKVSSSSFRFVASKTAINESSVAIDVEDVAYKRRAGFSHCADPPCYRG  
TNLDFGSCVOSRTIEEGYRTDTAGNVAVSIPTNPLVNAVSPYVAEYNEMNRSLI  
RRCCYACVCLVSVLLIMSGLAIFI"  
11843..11998  
/note="ORF4"  
/codon\_start=1  
/product="p6 protein"  
/protein\_id="CA77164.1"  
/db\_xref="PID:e1375377"  
/db\_xref="PID:94239718"  
/db\_xref="GI:4239718"  
/translation="MDCVIGGFLFVLGIAVSAFAGLIIVITITRYCTKPVNTSP  
YGHATV"  
12004..13788  
/note="ORF5"  
/codon\_start=1  
/product="p65 protein"  
/protein\_id="CA77165.1"  
/db\_xref="PID:e1375378"  
/db\_xref="PID:94239719"  
/db\_xref="GI:4239719"  
/translation="WVLIGDGTFTFSVAMATSELYILKSNSSYIPCLFLHADP  
NSVSGYDAEVLASGEPSFYKDKRWGCTAKRYQTLHRKSPRYTEKRGTKS  
VVPPLSLNDLGLNALPLIASYAKSILSDARVNVSCGTGVCSPAGYNTLOR  
AFQOISMSGSCYIINEPSAAVSTLPKLSADKILAVDFGGTFDVSIVSRL  
PTFVRSRGDMNGRDIDKRLSDKIEYEMADFIPOKELAVSLKEALSLQTDYKAT  
VTHHMSRTISIDOTALREIAVSFINRTIDILIOYKVSMSPESSKLTVVGGSYL  
PGLDMLATVPFGIVPEADARFARGACALYSCLDGRKALLIDCITHHLSYTP  
SADSVYAAAGSPIPEEGOKILRKCYSTNQRMFEGDYEFERNERTYASVSL  
FTLGVMHVPNDVENTLVTKVDSMGKVEFYLGPSELVNWGTSRYDTGMPPTNR  
LVKLSDVNNAALVALTLTRERKREKFLRTLFDTLADLRRTASLSEYSKYPITR  
NDIDVSSRMGIIVVSKVLRGSLERLPL"  
13712..15319  
/note="ORF6"  
/codon\_start=1  
/product="p61 protein"  
/protein\_id="CA77166.1"  
/db\_xref="PID:e1375379"  
/db\_xref="PID:94239720"  
/db\_xref="GI:4239720"  
/translation="MSHHWGSLEFRKFGAIMEKYLSESTRSDERNVSLDHTLSS  
GVVVRKOSLNAPOGTFENELALINSVINDVELTGMPLKSLMTGIEDKVPDELI  
SVDPHEVGCRTLNDVESYLISRGESADVAVEHSMCLNSCGRLSTEIDAYKTL  
VETKNFDSNVSGVITKLELYLSYCISLYKHKMDKNDFNLITPFNCLVLAISGL  
EYERKTDNPLTLGLIEFLCNKYYSFEKYNLDNRLFKSVLPVYLVMDISEPD  
PYDERVLIPDPDPLVDLPKLNIDHMVVVGNQIROLEVYESDALDDLSOHDIRL  
AADNDLAVDLRMAGMYYCYRCAIDRYVERPTLFLPKLLRQDDGESCLRMSA  
VEALENLVQVNRDINVRQFMGRHSKVALRLNGLRFPPISSVRLPVHGLIYVD  
FYKRVDDAVTADELESRLRSSVDVCMKDRVSIPTPPFNLRGSSRTFRGRGARG  
ASSRMSRDVATSGFNLPYHGLYDTS"  
15294..16016  
/note="ORF7"  
/codon\_start=1  
/product="p27 protein"  
/protein\_id="CA77167.1"  
/db\_xref="PID:e1375380"  
/db\_xref="PID:94239721"  
/db\_xref="GI:4239721"  
/translation="MAGYITLDPNTDDKEMDVSAAVPGKYPDVIEKTVANRSDALIE  
GVISKLDINSIYEDTEKFTGEHLKYVATMDAFLLENYKTKTEDLVHLAMIQRLC  
ITSTKTRKFDKGCISYVGGSRKLMKDVFPPIISKTHRETPNALRKYACTFEE  
LHLCMARLPDIYENKRTTKAGTPHLKGYLSADFLSGSLPGYSEHERGIIIRASESML  
ARROGEATEELNLRDLGKYL"  
16116..16787  
/note="ORF7"

/codon\_start=1  
/product="p25 protein"  
/protein\_id="CA77168.1"  
/db\_xref="PID:e1375381"  
/db\_xref="PID:94239722"  
/db\_xref="GI:4239722"  
/translation="MDEETKLLKRNKEKKEGGDYYVAESSFESVNIHIDPTLITMD  
VRLSTQONALNRDLFLALGKYPNLPDKDQFHLAMLYRLAVSSSLQSDDDTGG  
ITVREGVEYDLSDKLMTDIYNSKIGINRTNALRVGRINDALYLAFCONRNLSTG  
GRPLDAGIPAGYHYLCADFLTAGLTDLECAVYIQAKEQLKRRGADEVVTVNRQIG  
KFNRR"  
16753..17256  
CDS

Query Match 78.1%; Score 16.4; DB 17; Length 19259;  
Best Local Similarity 94.4%; Pred. No. 2.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 4 gataataatggagcgt 21  
Db 10614 GATAAATATTGACCGT 10597  
|||||  
|||||

Search completed: August 18, 1999, 17:17:40  
Job time: 3506 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 17:18:36 : Search time 148.09 seconds  
(without alignments)  
35.479 Million cell updates/sec

Title: US-09-004-395-5

Perfect score: 21

Sequence: 1 tgggataaataatgagcgt 21

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	21	100.0	110000	1	X20248_01	Continuation (2 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide)
C 2	21	100.0	110000	1	X20248_02	Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide)
C 3	15.8	75.2	1088	1	N90126	CDNA encoding huma
C 4	15.8	75.2	1193	1	O04664	Encodes alpha subu
C 5	15.8	75.2	1199	1	O14736	Human Fc(gamma)R
C 6	15.8	75.2	1150	1	O34840	Human high affinity
C 7	15.8	75.2	1174	1	O51020	Human Fc(gamma)R
C 8	15.8	75.2	1829	1	T45041	Helicobacter CP2 a
C 9	15.8	75.2	1803	1	T45042	Helicobacter CP2 a
C 10	15.8	75.2	1810	1	T45043	Helicobacter CP2 a
C 11	15.8	75.2	1810	1	T85615	Helicobacter CP2 a
C 12	15.8	75.2	1518	1	V20911	Alpha subunit of h
C 13	15.8	75.2	1518	1	V20910	Helicobacter pylori
C 14	15.8	75.2	773	1	V20402	Human IGE receptor
C 15	15.8	75.2	1198	1	V36343	CDNA encoding the
C 16	15.8	75.2	1780	1	X14142	H. pylori GHPD 358
C 17	15.8	75.2	968	1	X20345	Borrelia burgdorferi
C 18	15.8	75.2	2462	1	X20299	Borrelia burgdorferi
C 19	15.8	75.2	1177	1	X20333	Borrelia burgdorferi
C 20	15.2	72.4	3427	1	O56733	H. pylori protein
C 21	15.2	72.4	831	1	T67692	H. pylori protein
C 22	15.2	72.4	894	1	T68071	H. pylori protein
C 23	15.2	72.4	894	1	V24748	H. pylori ORF 11ce
C 24	15.2	72.4	735	1	V25016	H. pylori cytoplasm
C 25	15.2	72.4	28690	1	X13075	Enterococcus faecalis
C 26	15.2	72.4	840	1	X14361	H. pylori GHPD 355
C 27	15.2	72.4	850	1	X14060	H. pylori GHPD 355
C 28	14.8	70.5	1641	1	T34671	Nuclear inclusion
C 29	14.8	70.5	1156	1	T34672	Coat protein gene
C 30	14.8	70.5	1228	1	T84170	DNA encoding a hlg
C 31	14.8	70.5	12368	1	V31206	E. coli J96 pathog
C 32	14.8	70.5	1575	1	V07655	CDNA encoding a hu
C 33	14.8	70.5	1342	1	V62449	Human neurotensin
C 34	14.8	70.5	474	1	V75233	Staphylococcus aur
C 35	14.8	70.5	13868	1	V74484	Staphylococcus aur
C 36	14.8	70.5	5663	1	X12989	Staphylococcus aur
C 37	14.8	70.5	963	1	X20142	Enterococcus faecalis
C 38	14.6	69.5	1801	1	V16947	Enterococcus faecalis
C 39	14.6	69.5	2619	1	V74173	Rubisco large subu
C 40	14.6	69.5	1801	1	X16778	Human hUBB3 gene
C 41	14.4	68.6	3467	1	O06550	Pea Rubisco large
C 42	14.4	68.6	110000	1	V21209_06	Fibrinogen-binding
C 43	14.4	68.6	13740	1	V74365	Staphylococcus aur

## ALIGNMENTS

44 14.4 68.6 15747 1 X13018 Enterococcus faecalis  
45 14.4 68.6 450 1 X40992 Human secreted pro

RESULT 1  
X20248\_01/c  
Continuation (2 of 10) of X20248 from base 100001 (Borrelia burgdorferi polynucleotide)

WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
WP Fragment Name Begin End  
WP X20248\_00 1 110000  
WP X20248\_01 100001 210000  
WP X20248\_02 200001 310000  
WP X20248\_03 300001 410000  
WP X20248\_04 400001 510000  
WP X20248\_05 500001 610000  
WP X20248\_06 600001 710000  
WP X20248\_07 700001 810000  
WP X20248\_08 800001 910000  
WP X20248\_09 900001 910715

Query Match  
Best Local Similarity 100.0%; Score 21; DB 1; Length 110000;  
Pred. No. 0.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tgggataaataatgagcgt 21  
DB 104904 TGGGATAAATAATGAGCGT 104884

RESULT 2  
X20248\_02/c  
Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide)

WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
WP Fragment Name Begin End  
WP X20248\_00 1 110000  
WP X20248\_01 100001 210000  
WP X20248\_02 200001 310000  
WP X20248\_03 300001 410000  
WP X20248\_04 400001 510000  
WP X20248\_05 500001 610000  
WP X20248\_06 600001 710000  
WP X20248\_07 700001 810000  
WP X20248\_08 800001 910000  
WP X20248\_09 900001 910715

Query Match  
Best Local Similarity 100.0%; Score 21; DB 1; Length 110000;  
Pred. No. 0.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tgggataaataatgagcgt 21  
DB 4904 TGGGATAAATAATGAGCGT 4884

RESULT 3  
N90126/c  
ID N90126 standard; cDNA; 1088 BP.  
AC N90126  
DT 1-NOV-1989 (first entry)  
DE cDNA encoding human mast cell IGE receptor alpha subunit  
KW cDNA; immunoglobulin E receptor alpha subunit; treats  
OS allergies; design non-peptide drugs; human.  
FH Homo sapiens (Human); mast cell line, KU812.  
FT Key Location/Qualifiers  
FT cds 35..805  
FT /\*tag= a  
PD W08905352-A.  
PD 15-JUN-1989.

PF 29-NOV-1988; U04255.  
 PR 01-DEC-1987; US-127214.  
 PA (HARD) Harvard College.  
 PI Siraganian R, Shlimizu A, Leder P, Benfey P;  
 DR WPI: 89-192698/26.  
 DR P-PSDB: P90385.  
 PT cDNA encoding IGE receptor alpha-subunit - used to treat allergies.  
 PS Disclosure; Fig 4; 18pp; English.  
 CC cDNA sequence encoding immunoglobulin E receptor alpha  
 CC subunit of human mast cell IGE surface receptor (see corresp. P90385).  
 CC used to produce antibodies which can diagnose IGE receptor levels,  
 CC measure and treat allergies, and design non-peptide drugs.  
 SQ Sequence 1088 BP; 347 A; 216 C; 229 G; 296 T;

Query Match  
 Best Local Similarity 75.2%; Score 15.8; DB 1; Length 1088;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tgggataataattggagc 19  
 |||||  
 DB 661 TGGGATATAAAATTGTACG 643

RESULT 4  
 ID 004644 standard; DNA; 1193 BP.  
 AC 004644;  
 DT 02-OCT-1990 (first entry)  
 DE Encodes alpha subunit of human high affinity IGE receptor  
 KW high affinity IGE receptor; alpha subunit of IGE receptor;  
 KW human; allergic response; ss.  
 OS synthetic.  
 PN WO9004640-A.  
 PD 03-MAY-1990.  
 PF 18-OCT-1988; U04628.  
 PR 18-OCT-1988; US-259065.  
 PA (USDC) US Sec of Commerce.  
 PI Kinet JP, Metzger H;  
 DR WPI: 90-164023/21.  
 DR P-PSDB: R05025.  
 PT DNA sequences for subunit peptides of high affinity IGE receptor -  
 PT and derived polypeptides, for therapy and diagnosis of  
 PT allergies, and studies of IGE receptor interaction  
 PS Disclosure; P; English.  
 CC The high affinity receptor is a tetrameric complex consisting of 2  
 CC gamma subunits and one each of subunits alpha and beta. It is  
 CC expressed on mast cells and is involved in the allergic response.  
 CC COS-7 cells cotransfected with cDNA for all 3 intact subunit types  
 CC (derived from rat basophilic leukaemia cells) express receptor on  
 CC their surfaces. Detailed study of the receptors is now possible.  
 CC See also Q04643 and Q04645-6.  
 SQ Sequence 1193 BP; 384 A; 241 C; 247 G; 321 T;

Query Match  
 Best Local Similarity 75.2%; Score 15.8; DB 1; Length 1193;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tgggataataattggagc 19  
 |||||  
 DB 733 TGGGATATAAAATTGTACG 715

RESULT 5  
 ID 014736 standard; cDNA; 1199 BP.  
 AC 014736;  
 DT 03-FEB-1992 (first entry)  
 DE Human Fc(epsilon)RI alpha cDNA.  
 KW Immunoglobulin; receptor; high affinity receptor; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers

FT cds 107.880  
 FT /\*tag= a  
 FT /product= alpha\_subunit  
 PN US7626704-A.  
 PD 15-OCT-1991.  
 PF 14-DEC-1990; 151091.  
 PR 14-DEC-1990; US-626704.  
 PA (USSH ) NAT INST OF HEALTH.  
 PI Kinet JP, Metzger H;  
 DR WPI: 91-346755/47.  
 DR P-PSDB: R14772.  
 PT DNA coding alpha, beta and gamma units of IGE high affinity  
 PT receptor - are used to prepare recombinant polypeptide(s) for  
 PT treating allergy; drug screening or monitoring IGE level.  
 PS Disclosure; Fig 1; 58pp; English.  
 CC A lambda gt10 library was prepared from poly-A RNA isolated from  
 CC Knu2 cells. Screening was by a cDNA fragment from the rat  
 CC Fc(epsilon)RI alpha cDNA corresponding to nucleotides 119-781.  
 CC Positive clones were subcloned and sequenced.  
 SQ Sequence 1199 BP; 390 A; 241 C; 247 G; 321 T;

Query Match  
 Best Local Similarity 75.2%; Score 15.8; DB 1; Length 1199;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tgggataataattggagc 19  
 |||||  
 DB 733 TGGGATATAAAATTGTACG 715

RESULT 6  
 ID 034840 standard; DNA; 1150 BP.  
 AC 034840;  
 DT 18-APR-1996 (revised)  
 DT 07-MAY-1993 (first entry)  
 DE Human high affinity IGE receptor alpha subunit gene.  
 KW Allergic response; antagonist; drug screening; vaccine; allergic;  
 KW reaction; diagnosis; therapy; FCER1; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 124.847  
 FT /\*tag= a  
 FT misc\_feature 601  
 FT /\*tag= b  
 FT /\*note= "sequence unreadable at this point"  
 FT 745  
 FT /\*tag= c  
 FT /\*note= "sequence unreadable at this point"  
 FT 817  
 FT /\*tag= d  
 FT /\*note= "sequence unreadable at this point"  
 PN US7547892-A.  
 PD 01-DEC-1992.  
 PF 02-JUL-1990; 230085.  
 PR 24-FEB-1988; US-160457.  
 PR 02-JUL-1990; US-547892.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PI Kinet JP;  
 DR WPI: 93-008901/01.  
 DR P-PSDB: R30483.  
 PT Novel DNA encoding alpha sub-unit of human high affinity IGE  
 PT receptor - useful as antagonist to prevent allergic response, as  
 PT reagent in drug screening and for antibody generation  
 PS Disclosure; Fig 1; 34pp; English.  
 CC The sequence is that of the gene encoding the human IGE high affinity  
 CC receptor alpha subunit (FCER1). It may be used to synthesise the  
 CC human FCER1 alpha subunit or to synthesise cDNA sequences to  
 CC construct DNA probes useful in diagnostic assays.  
 SQ Sequence 1150 BP; 373 A; 237 C; 227 G; 310 T;

Query Match 75.2%; Score 15.8; DB 1; Length 1150;  
 Best Local Similarity 89.5%; Pred. No. 44;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tgggataaataatggagc 19  
 |||||  
 Db 706 TGGGATRAAAAATTGTAGC 688

RESULT 7  
 051020/c  
 ID 051020 standard; DNA: 1174 BP.  
 AC 051020;  
 DT 21-JUN-1994 (first entry)  
 DE Human FCERI alpha gene.  
 KW 1GE; immunoglobulin E receptor; beta subunit; basophilic allergy;  
 KM aggregation; signal transduction; diagnosis; antagonist; ss.  
 OS Homo sapiens.  
 FH Key  
 FT cds Location/Qualifiers  
 FT 107..881  
 FT /\*tag- a  
 PN WO9321317-A.  
 PD 28-OCT-1993.  
 PF 16-APR-1993; U03419.  
 PR 16-APR-1992; US-869933.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PI Kinet JP;  
 DR WPI: 93-351727/44.  
 P-PSDB: R42336.  
 CC Immunoglobulin E receptor human beta sub-unit isolation - using  
 PT 1st strand reverse transcripts from human basophils as templates  
 PS for a polymerase chain reaction, used to treat and diagnose  
 PT allergic diseases  
 CC Claim 1: Fig 1: 154pp: English.  
 CC The sequence is that of the human FCRI alpha subunit, isolated by  
 CC using first strand reverse transcriptase from human basophils by PCR.  
 CC The gene and its prod. can be used to identify human beta subunit  
 CC FCERI inhibitors (immunoglobulin E receptor) which inhibit the  
 CC binding of IGE to its receptor and inhibit the aggregation function  
 CC of the receptor or the signal transducing function related to  
 CC allergic response. Such inhibitors can be used for the treatment or  
 CC prevention of allergic disease.  
 CC See also Q51021-4.  
 SO Sequence 1174 BP; 378 A; 240 C; 241 G; 315 T;

Query Match 75.2%; Score 15.8; DB 1; Length 1174;  
 Best Local Similarity 89.5%; Pred. No. 44;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 tgggataaataatggagc 19  
 |||||  
 Db 733 TGGGATRAAAAATTGTAGC 715

RESULT 8  
 T45041  
 ID T45041 standard; DNA: 1829 BP.  
 AC T45041;  
 DT 24-FEB-1997 (first entry)  
 DE Helicobacter CP2 antigen structural gene.  
 KW CP2; antigen; vaccine; diagnosis; gastritis; ds.  
 OS Helicobacter pylori strain ATCC 43504.  
 FH Key  
 FT cds Location/Qualifiers  
 FT 21..1538  
 FT /\*tag- a  
 FT complement (18..38)  
 FT /\*tag- b  
 FT /note- "forward primer for CP2 DNA amplification"  
 FT complement (97..118)  
 FT /\*tag- c  
 FT /note- "forward primer for H. pylori detection"  
 FT primer\_bind 364..385

FT /\*tag- d  
 FT /note- "reverse primer for H. pylori detection"  
 FT complement (841..863)  
 FT /\*tag- e  
 FT /note- "forward primer for H. pylori detection"  
 FT 1261..1283  
 FT /\*tag- f  
 FT /note- "reverse primer for H. pylori detection"  
 FT 1800..1820  
 FT /\*tag- g  
 FT /note- "reverse primer for CP2 DNA amplification"

EP-745674-A2.  
 PD 04-DEC-1996.  
 PF 30-MAY-1996; 108637.  
 PR 02-JUN-1995; JP-136564.  
 PR 05-APR-1996; JP-083512.  
 PA (WAKP ) WAKO PURE CHEM IND LTD.  
 PI Hirayasu K, Kawabata T, Sugiyama T, Tanaka T;  
 DR WPI: 97-013697/02.  
 P-PSDB: W06481.  
 CC Helicobacter pylori CP2 antigen structural polypeptide(s) - also  
 PT corresponding DNA, useful as reagents for detecting H. pylori and in  
 PT the production of vaccines  
 CC Claim 3: Page 16-18; 29pp: English.  
 CC A structural gene (T45041) of Helicobacter pylori ATCC 43504 codes  
 CC for the CP2 antigen (W06481) useful in vaccine prodn. Genomic  
 CC fragments of H. pylori DNA were inserted into vector pBluescript,  
 CC and E. coli JM109 transformants were screened using a probe based  
 CC on the N-terminal sequence of CP2. PCR primers (T45048-49) based on  
 CC isolated clones were then used to amplify the structural gene. The  
 CC gene sequence differs slightly from those of CP2 genes (see also  
 CC T45042 and T45043) obtd. by PCR amplification. CP2 genes can be  
 CC used to produce recombinant CP2 in transformed host cells, or to  
 CC design primers (see also T45044-47) useful in the specific  
 CC detection of H. pylori.  
 SO Sequence 1829 BP; 589 A; 337 C; 409 G; 494 T;

Query Match 75.2%; Score 15.8; DB 1; Length 1829;  
 Best Local Similarity 89.5%; Pred. No. 45;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 ggaataaataatggagc 21  
 |||||  
 Db 1403 GGATTAACAATTGAGCAT 1421

RESULT 9  
 T45042  
 ID T45042 standard; DNA: 1803 BP.  
 AC T45042;  
 DT 24-FEB-1997 (first entry)  
 DE Helicobacter CP2 antigen structural gene CP2-PCR1.  
 KW CP2; antigen; vaccine; diagnosis; gastritis; ds.  
 OS Helicobacter pylori strain ATCC 43504.  
 FH Key  
 FT cds Location/Qualifiers  
 FT 4..1521  
 FT /\*tag- a  
 FT complement (80..101)  
 FT /\*tag- b  
 FT /note- "forward primer for H. pylori detection"  
 FT 347..368  
 FT /\*tag- c  
 FT /note- "reverse primer for H. pylori detection"  
 FT complement (824..846)  
 FT /\*tag- d  
 FT /note- "forward primer for H. pylori detection"  
 FT 1244..1266  
 FT /\*tag- e  
 FT /note- "reverse primer for H. pylori detection"  
 FT primer\_bind BP-745674-A2.  
 PD 04-DEC-1996.  
 PF 30-MAY-1996; 108637.





RESULT 12  
ID V20911 standard; cDNA: 1518 BP.  
AC V20911:  
DE 20-JUL-1998 (first entry)  
DE Helicobacter pylori isolate 921023 catalase gene.  
KM Treatment; prevention; gastroduodenal disease; detection; ss.  
OS Helicobacter pylori.  
FH Key Location/Qualifiers  
FT 1..1518  
FT CDS /\*tag= a  
FT /product= catalase  
FN WO9806853-A1.  
PD 19-FEB-1998.  
PE 14-AUG-1997: AU0515.  
PR 15-AUG-1996: US-695987.  
PA (CSLC-) CSL LTD.  
PY (UYNE-) UNIV NEW SOUTH WALES.  
PI Doldge CV, Hazell SL, Lee A, Radcliff FJ;  
DR WPI: 98-159544/14.  
P-PSDB: W51146.  
PT New isolated Helicobacter catalase nucleic acid - used to develop  
PT products for the treatment or prevention of Helicobacter infection,  
PT particularly H. pylori gastroduodenal disease  
PS Claim 29; Page 33-37; 46pp: English.  
CC The Helicobacter pylori catalase gene encodes a protein of 58650 Da.  
CC The protein, gene sequence and products, such as an immunogenic fragment,  
CC particularly H. pylori infections which cause gastroduodenal disease,  
CC They can also be used for the detection and diagnosis of Helicobacter  
CC infection.  
SQ Sequence 1518 BP; 487 A; 289 C; 350 G; 392 T;  
  
Query Match 75.2%; Score 15.8; DB 1; Length 1518;  
Best Local Similarity 89.5%; Pred. No. 45;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
OY 3 ggataataattggagcgt 21  
DB 1383 GGATTAACAATTGGAGCAT 1401  
||||| |||||||  
RESULT 13  
ID V20910 standard; cDNA: 1518 BP.  
AC V20910:  
DE 20-JUL-1998 (first entry)  
DE Helicobacter pylori isolate RUI catalase gene.  
KM Treatment; prevention; gastroduodenal disease; detection; ss.  
OS Helicobacter pylori.  
FH Key Location/Qualifiers  
FT 1..1518  
FT CDS /\*tag= a  
FT /product= catalase  
FN WO9806853-A1.  
PD 19-FEB-1998.  
PE 14-AUG-1997: AU0515.  
PR 15-AUG-1996: US-695987.  
PA (CSLC-) CSL LTD.  
PY (UYNE-) UNIV NEW SOUTH WALES.  
PI Doldge CV, Hazell SL, Lee A, Radcliff FJ;  
DR WPI: 98-159544/14.  
P-PSDB: W52810.  
PT New isolated Helicobacter catalase nucleic acid - used to develop  
PT products for the treatment or prevention of Helicobacter infection,  
PT particularly H. pylori gastroduodenal disease  
PS Claim 29; Page 31-33; 46pp: English.  
CC The Helicobacter pylori catalase gene encodes a protein of 58650 Da.  
CC The protein, gene sequence and products, such as an immunogenic fragment,  
CC can be used for the treatment or prevention of Helicobacter infection,  
CC particularly H. pylori infections which cause gastroduodenal disease.

CC They can also be used for the detection and diagnosis of Helicobacter  
CC infection.  
SQ Sequence 1518 BP; 489 A; 295 C; 349 G; 385 T;  
  
Query Match 75.2%; Score 15.8; DB 1; Length 1518;  
Best Local Similarity 89.5%; Pred. No. 45;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
OY 3 ggataataattggagcgt 21  
DB 1383 GGATTAACAATTGGAGCAT 1401  
||||| |||||||  
RESULT 14  
ID V20402 standard; cDNA: 773 BP.  
AC V20402:  
DE 20-JUL-1998 (first entry)  
DE Human IGE receptor Fc-epsilon-RI alpha chain DNA.  
KM Fc-epsilon-RI alpha chain; IGE receptor; human serum albumin;  
KM fusion protein; allergy; atopic dermatitis; asthma; urticaria;  
KM hay fever; eczema; anaphylaxis; gene therapy; diagnosis;  
KM Transgenic animal; ds.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT sig\_peptide 1..75  
FT mat\_peptide 76..773  
FT /\*tag= a  
FT /\*tag= b  
FN WO9804718-A1.  
PD 05-FEB-1998.  
PE 25-JUL-1997: E04066.  
PR 26-JUL-1996: US-690216.  
PA (NOVS) NOVARTIS AG.  
PI Digan ME, Gram H, Lake P;  
DR WPI: 98-130705/12.  
P-PSDB: W48094.  
PT New fusion polypeptide for, e.g. diagnosing allergies - comprises  
PT immunoglobulin F-binding domain fused to human serum albumin  
PS Disclosure; Page 56; 77pp: English.  
CC This nucleotide sequence codes the dominant form of full-length  
CC native human IGE receptor Fc-epsilon-RI alpha chain (see W48094).  
CC A claimed fusion protein (FP) comprises an IGE binding domain fused  
CC to at least one human serum albumin (HSA) component (see W38095),  
CC optionally via a peptide linker, and is especially a dimeric FP  
CC (see W48096) comprising HSA fused, at each of its N- and C-termini,  
CC to the extracellular domain (i.e. mature protein) of Fc-epsilon-RI  
CC alpha chain. Also claimed are: nucleic acids encoding the FP; a  
CC vector; a process for preparing the FP; a method of performing gene  
CC therapy in humans that comprises removing somatic cells from a  
CC patient, genetically modifying them in culture by insertion of a  
CC polynucleotide that encodes the FP, and reintroducing the modified  
CC cells into the patient so that the FP is expressed by the cells of  
CC the patient; and use of the FP in an in vitro diagnostic assay to  
CC determine the level of IGE or auto-antibodies to Fc epsilon RI in a  
CC sample. The products can be used in the prevention and/or treatment  
CC of IGE-mediated allergic diseases and related disorders such as  
CC atopic dermatitis, atopic asthma, chronic urticaria, hay fever and  
CC eczema. Compared with using IGE binding domain alone, the FP has a  
CC longer serum life, and thus greater activity, without a loss of  
CC ability to bind serum IGE or circulating auto-antibodies.  
SQ Sequence 773 BP; 220 A; 173 C; 176 G; 204 T;  
  
Query Match 75.2%; Score 15.8; DB 1; Length 773;  
Best Local Similarity 89.5%; Pred. No. 43;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
OY 1 ttggataataattggagc 19  
DB 627 TGGGATTAATAAATTGTAGC 609  
||||| |||||||

RESULT 15

V36343/c standard; CDNA: 1198 BP.

ID V36343:

DT 07-OCT-1998 (first entry)

DE CDNA encoding the alpha chain of a Fc epsilon receptor.

KW Alpha chain; human; Fc epsilon receptor; canine; equine; feline;

KW immunoglobulin E; IGE; detection; diagnose; allergy; atopic disease;

KW hyper-IGE syndrome; internal parasitic infection; B cell neoplasia;

KW flea allergy; heartworm infection; ds.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 107..880

FI /\*tag= a

PN WO9823964-A1.

PD 04-JUN-1998.

PF 24-NOV-1997; U21651.

PR 26-NOV-1996; US-756387.

PA (HESK-) HESKA CORP.

PI Frank RG, Porter JP, Rushlow KE, Wassom DL;

DR WPI: 98-322873/28.

DR P-PSDB: W61190.

PT Detection of non-human immunoglobulin E - by complex formation with

PT human Fc epsilon receptor, used for, e.g. diagnosis of allergy and

PS Claim 11: Pages 36-37; 70pp; English.

CC The present sequence encodes the alpha chain of the human Fc epsilon

CC receptor. Detection of canine, equine or feline immunoglobulin E (IGE)

CC comprises reacting isolated human Fc epsilon receptor with the test

CC sample and detecting formation of a IGE-receptor complex. Detection of

CC IGE is used to diagnose allergy, atopic disease, hyper-IGE syndrome,

CC internal parasitic infections or B cell neoplasia, and for measuring

CC effect of treatments. Most particularly flea allergy in dogs and cats

CC is detected, and also heartworm infection.

SQ Sequence 1198 BP; 389 A; 243 C; 245 G; 321 T;

Query Match 75.2%; Score 15.8; DB 1; Length 1198;

Best Local Similarity 89.5%; Pred. No. 44;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ttggataaataatgtgagc 19

Db 733 TCGGATATAAAATTGTAGC 715

Search completed: August 18, 1999, 17:18:42  
Job time: 3274 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 17:04:46 ; Search time 767.16 seconds  
(without alignments)  
53.996 Million cell updates/sec

Title: US-09-004-395-5

Perfect score: 21

Sequence: 1 tgggataaataatgagcgt 21

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	17.4	82.9	308	25	D81091	D81091 HUM129E12B
C 2	16.2	77.1	531	27	C06497	C06497 C06497 Huma
C 3	16.2	77.1	499	40	AA962861	AA962861 U30-487T3
C 4	16.2	77.1	414	46	AI460669	AI460669 sa71909.y
C 5	16.2	77.1	417	46	AI461277	AI461277 sa60H09.y
C 6	16.2	77.1	482	47	AI522878	AI522878 sa74H08.y
C 7	15.8	75.2	354	30	AA226025	AA226025 nc09e10.s
C 8	15.8	75.2	370	47	AI511010	AI511010 UI-R-C2P-
C 9	15.8	75.2	446	50	AI676097	AI676097 wC05C12.x
C 10	15.8	75.2	540	50	AI685796	AI685796 tu20908.x
C 11	15.4	73.3	554	26	W92119	W92119 zh48a09.r1
C 12	15.4	73.3	421	35	AA574291	AA574291 nf45a09.s
C 13	15.4	73.3	375	35	AA574359	AA574359 nf45a09.s
C 14	15.4	73.3	438	43	AI218562	AI218562 qh19e01.x
C 15	15.4	73.3	416	45	AI350781	AI350781 q921d04.x
C 16	15.4	73.3	535	48	AI545362	AI545362 fb74c12.y
C 17	15.4	73.3	479	50	AI688192	AI688192 wC87a04.x
C 18	15.4	73.3	219	50	AV030487	AV030487 AV030487
C 19	15.4	73.3	456	51	AI722228	AI722228 fd20f09.y
C 20	15.2	72.4	367	20	T18656	T18656 SC01A07-T7
C 21	15.2	72.4	450	20	T23327	T23327 SC07g04 mem
C 22	15.2	72.4	420	22	H14450	H14450 Y125d03.r1
C 23	15.2	72.4	448	22	R54680	R54680 Y174f09.r1
C 24	15.2	72.4	391	23	H24807	H24807 Y142c07.r1
C 25	15.2	72.4	443	24	N33607	N33607 Y22c12.s1
C 26	15.2	72.4	553	25	N98431	N98431 zb77e11.s1
C 27	15.2	72.4	403	29	AA186829	AA186829 zp73h10.r
C 28	15.2	72.4	447	34	AA454469	AA454469 zw28c07.r
C 29	15.2	72.4	382	34	AA481687	AA481687 zw45c03.r
C 30	15.2	72.4	300	35	C38476	C38476 C38476 YUJ1
C 31	15.2	72.4	300	35	C51836	C51836 C51836 YUJ1
C 32	15.2	72.4	193	39	AA894898	AA894898 n247h11.s
C 33	15.2	72.4	480	40	AA976721	AA976721 Oq06e04.s
C 34	15.2	72.4	368	40	C82644	C82644 C82644 rabd
C 35	15.2	72.4	368	40	C83500	C83500 C83500 rabd
C 36	15.2	72.4	342	42	AI150563	AI150563 gf36e04.x
C 37	15.2	72.4	447	44	AI251447	AI251447 q774g06.x
C 38	15.2	72.4	603	44	AI262642	AI262642 qK34f04.x
C 39	15.2	72.4	422	45	AI341339	AI341339 qX90h10.x
C 40	15.2	72.4	232	50	AV043793	AV043793 AV043793
C 41	15.2	72.4	272	50	AV046241	AV046241 AV046241
C 42	15.2	72.4	432	51	AI733829	AI733829 qK34f04.x
C 43	15.2	72.4	656	54	HSM011851	HSM011851 Homo_sapi
C 44	15.2	72.4	556	54	AI231334	AI231334 EST228022
C 45	14.8	70.5	876	50	AU066662	AU066662 AU066662

#### ALIGNMENTS

RESULT 1  
D81091 308 bp mRNA EST  
LOCUS HUM129E12B Human fetal brain (Tfujlwara) Homo sapiens CDNA clone  
DEFINITION GEN-129E12 5', mRNA sequence.  
ACCESSION D81091  
NID g1178968  
VERSION D81091.1 GI:1178968

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS 1 (bases 1 to 308)  
Fujimura, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takachi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H., Shin, S. and Nakamura, Y.  
TITLE Fujimura et al. (1995)  
JOURNAL Unpublished (1995)  
COMMENT

FEATURES  
source 1..308  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Human fetal brain (TFujimura)"  
/dev\_stage="fetal"  
/note="Organ: brain"

BASE COUNT 91 a 65 c 55 g 72 t 25 others  
ORIGIN

Query Match 82.9%; Score 17.4; DB 25; Length 308;  
Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgggataataattggagcgt 21  
|||||  
Db 157 TGGGATTAATATYTGAGAGT 137

RESULT 2  
LOCUS C06497 531 bp mRNA EST 16-OCT-1996  
DEFINITION C06497 Human pancreatic islet Homo sapiens cDNA clone hbc5859 similar to cysteine protease CMH-1, mRNA sequence.  
ACCESSION C06497  
NID Q1503273  
VERSION C06497.1 GI:1503273  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 531)  
Takeda, J.  
TITLE Human pancreatic islet ESTs  
JOURNAL Unpublished (1995)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:315339.

FEATURES  
source 1..531  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="hbc5859"  
/clone\_lib="Human pancreatic islet"  
/note="Vector: Lambda ZAPII; Site\_1: Eco RI; Site\_2: Xho I; mRNA was prepared from normal adult human islets. cDNA

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS 1 (bases 1 to 308)  
Fujimura, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takachi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H., Shin, S. and Nakamura, Y.  
TITLE Fujimura et al. (1995)  
JOURNAL Unpublished (1995)  
COMMENT

FEATURES  
source 1..308  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Human fetal brain (TFujimura)"  
/dev\_stage="fetal"  
/note="Organ: brain"

BASE COUNT 91 a 65 c 55 g 72 t 25 others  
ORIGIN

Query Match 82.9%; Score 17.4; DB 25; Length 308;  
Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgggataataattggagcgt 21  
|||||  
Db 157 TGGGATTAATATYTGAGAGT 137

RESULT 2  
LOCUS C06497 531 bp mRNA EST 16-OCT-1996  
DEFINITION C06497 Human pancreatic islet Homo sapiens cDNA clone hbc5859 similar to cysteine protease CMH-1, mRNA sequence.  
ACCESSION C06497  
NID Q1503273  
VERSION C06497.1 GI:1503273  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 531)  
Takeda, J.  
TITLE Human pancreatic islet ESTs  
JOURNAL Unpublished (1995)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:315339.

FEATURES  
source 1..531  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="hbc5859"  
/clone\_lib="Human pancreatic islet"  
/note="Vector: Lambda ZAPII; Site\_1: Eco RI; Site\_2: Xho I; mRNA was prepared from normal adult human islets. cDNA

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS 1 (bases 1 to 308)  
Fujimura, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takachi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H., Shin, S. and Nakamura, Y.  
TITLE Fujimura et al. (1995)  
JOURNAL Unpublished (1995)  
COMMENT

FEATURES  
source 1..308  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Human fetal brain (TFujimura)"  
/dev\_stage="fetal"  
/note="Organ: brain"

BASE COUNT 91 a 65 c 55 g 72 t 25 others  
ORIGIN

Query Match 82.9%; Score 17.4; DB 25; Length 308;  
Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgggataataattggagcgt 21  
|||||  
Db 157 TGGGATTAATATYTGAGAGT 137

RESULT 2  
LOCUS C06497 531 bp mRNA EST 16-OCT-1996  
DEFINITION C06497 Human pancreatic islet Homo sapiens cDNA clone hbc5859 similar to cysteine protease CMH-1, mRNA sequence.  
ACCESSION C06497  
NID Q1503273  
VERSION C06497.1 GI:1503273  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 531)  
Takeda, J.  
TITLE Human pancreatic islet ESTs  
JOURNAL Unpublished (1995)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:315339.

FEATURES  
source 1..531  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="hbc5859"  
/clone\_lib="Human pancreatic islet"  
/note="Vector: Lambda ZAPII; Site\_1: Eco RI; Site\_2: Xho I; mRNA was prepared from normal adult human islets. cDNA

DB 447 TTGCATTAACATTGGAGCAT 427

RESULT 4  
AI460669/c  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AI460669 414 bp mRNA EST 09-MAR-1999  
sa11909.y1 Gm-c1004 Glycine max CDNA clone GENOME SYSTEMS CLONE ID:  
Gm-c1004-4793 5' similar to TR:O23807 O23807 LEDI-3 PROTEIN. ;  
mRNA sequence.  
AI460669  
94133550  
AI460669.1 GI:4313550  
EST.  
soybean.  
Glycine max  
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
Glycine.  
1 (bases 1 to 414)  
Shoemaker, R., Briner, D., Marra, M., Hillier, L., Kucaba, T.,  
Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M.,  
Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T.,  
Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S.,  
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
Wilson, R.  
Public Soybean EST Project  
Unpublished (1999)  
On Jun 5, 1998 this sequence version replaced gi:1189594.

TITLE  
JOURNAL  
COMMENT

REFERENCE  
AUTHORS

Contact: Shoemaker R./Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134 For further information  
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or  
info@genomesystems.com web site: www.genomesystems.com  
Seq primer: -40RP from Glibco  
High quality sequence stop: 392.  
Location/Qualifiers

FEATURES  
SOURCE

1..414  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1004-4793"  
/clone\_11b="Gm-c1004"  
/tissue\_type="root"  
/lab\_host="XL10-Gold"  
/note="Vector: pBluescript II XR; Site\_1: EcoRI; Site\_2:  
XhoI; Root CDNA. The mRNA was isolated from entire roots  
of 8 day old 'Williams' seedlings which were propagated on  
paper towels with distilled water. Stragene's cDNA  
Synthesis Kit (catalog #200401) was used to synthesize the  
cDNA. First-strand synthesis was performed with 5-methyl  
dCTP, hence the ligated cDNA is hemimethylated.  
Stragene's first-strand synthesis primer was used  
[GAGAGAGAGAGAGAGACTAGTCTCAG(T)-18]. After  
second-strand synthesis, the cDNA ends were 'polished'  
with clone Pfu DNA polymerase, ligated to EcoRI adapters,  
and phosphorylated. The XhoI site within the first-strand  
synthesis primer was restricted by digestion with XhoI;  
all XhoI sites in the cDNA would be protected by their  
hemimethylated status. The cDNA constructs were  
size-fractionated with a 500bp cutoff, using GlibcoBRL life  
technologies' cDNA size fractionation column. The column  
eluent was then ligated into Stragene's pBluescript II  
XR predigested vector (pBluescript II SK(+)) that had been  
digested with EcoRI and XhoI, and phosphorylated). Both  
the white and blue colonies appear to contain recombinant

plasmids with cDNA inserts. Blue colonies (9n-15) have been  
sequenced, and possess putative cDNA inserts. This library  
was constructed by Dr. Paul Keim & Virginia H. Coryell,  
Department of Biology, Box5640, Northern Arizona  
University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.  
Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax:  
520-523-7500, email: paul.keim@na.u.edu,  
virginia.coryell@na.u.edu"

BASE COUNT 116 a 79 c 108 g 111 t  
ORIGIN

Query Match 77.1% Score 16.2; DB 46; Length 414;  
Best Local Similarity 85.7%; Pred. No. 3.7e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tgggtaataattggagcgt 21  
||||| |||||||||  
Db 242 TGGGATTAATATTGGAGCAT 222

RESULT 5  
AI461277/c  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AI461277 417 bp mRNA EST 09-MAR-1999  
sa60h09.y1 Gm-c1004 Glycine max CDNA clone GENOME SYSTEMS CLONE ID:  
Gm-c1004-3738 5' similar to SW:P25\_SCHPO P30821 P25 PROTEIN ; , mRNA  
sequence.  
AI461277  
94314158  
AI461277.1 GI:4314158  
EST.  
soybean.  
Glycine max  
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
Glycine.  
1 (bases 1 to 417)  
Shoemaker, R., Briner, D., Marra, M., Hillier, L., Kucaba, T.,  
Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M.,  
Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T.,  
Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S.,  
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
Wilson, R.  
Public Soybean EST Project  
Unpublished (1999)  
On Jun 22, 1998 this sequence version replaced gi:3247126.

REFERENCE  
AUTHORS

Contact: Shoemaker R./Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134 For further information  
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or  
info@genomesystems.com web site: www.genomesystems.com  
Seq primer: -40RP from Glibco.  
Location/Qualifiers

FEATURES  
SOURCE

1..417  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/map="14q24.3"  
/clone="GENOME SYSTEMS CLONE ID: Gm-c1004-3738"  
/clone\_11b="Gm-c1004"  
/tissue\_type="root"  
/lab\_host="XL10-Gold"  
/note="Vector: pBluescript II XR; Site\_1: EcoRI; Site\_2:  
XhoI; Root CDNA. The mRNA was isolated from entire roots  
of 8 day old 'Williams' seedlings which were propagated on



TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1998 this sequence version replaced g1:1316602.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbp/image/image.html

FEATURES  
source Seq primer: -41m3 fwd. ET from Amersham.  
Location/Qualifiers

1. 354  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="864B10: 9: 9q22.2-q22.32: 9q22.32-q31.2"  
/clone="IMAGE:1007658"  
/clone\_lib="NCI-CGAP\_Prl"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="vector: PAM10: Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected, histologically normal  
prostate epithelial cells. Double-stranded cDNA was  
ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
cDNA with an adaptor-specific primer, and the resulting  
PCR product subcloned into PAM10 by the UDG-cloning  
method (Life Technologies). Average insert size is 600  
bp. NOTE: Not directionally cloned. This library was  
constructed by David Krizman."

BASE COUNT 94 a 95 c 37 g 128 t  
ORIGIN  
Query Match 75.2%; Score 15.8; DB 30; Length 354;  
Best Local Similarity 89.5%; Pred. No. 5.9e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgggataaataatgagc 19  
|||||  
DB 275 TCGGATAATGAGC 257  
RESULT 8  
A1511010 370 bp mRNA EST 15-MAR-1999  
LOCUS A1511010/c  
DEFINITION UI-R-C2P-ny-h-06-0-UI-s1 UI-R-C2P Rattus norvegicus cDNA clone  
ACCESSION A1511010  
NID 94416709  
VERSION A1511010.1 GI:4416709  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 370)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT On Jun 22, 1998 this sequence version replaced g1:3246630.

Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. CDNA Library Preparation: M.B. Soares Lab Clone distribution:  
clones will be available through Research Genetics (www.resgen.com)  
Seq primer: M13 forward.

FEATURES  
source Location/Qualifiers

1. 370  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-C2P-ny-h-06-0-UI"  
/clone\_lib="UI-R-C2P"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="vector: pTT3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-C2P  
library is a subtracted library derived from the UI-R-C1  
library, which is a subtracted library derived from the  
UI-R-C0 library. The UI-R-C0 library consisted of a  
mixture of individually tagged normalized libraries  
constructed from rat placenta, adult lung, brain, liver,  
kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day  
embryo. The tag is a string of 3-5 nucleotides present  
between the Not I site and the oligo-dT track which allows  
identification of the library of origin of a clone within  
the mixture. The subtracted library (UI-R-C2P) was  
constructed as follows: PCR amplified cDNA inserts from  
UI-R-C1 clones from which 3' ESTs had been derived was  
used as a driver in a hybridization with the UI-R-C1  
library in the form of single-stranded circles. The  
remaining single-stranded circles (subtracted library) was  
purified by hydroxyapatite column chromatography,  
converted to double-stranded circles and electroporated  
into DH10B bacteria (Life Technologies) to generate the  
UI-R-C2P library. This procedure has been previously  
described (Bonaldo, Lennon and Soares, Genome Research 6:  
791-806, 1996)."

BASE COUNT 105 a 80 c 50 g 134 t 1 others  
ORIGIN  
Query Match 75.2%; Score 15.8; DB 47; Length 370;  
Best Local Similarity 89.5%; Pred. No. 5.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gggataaataatgagcg 20  
|||||  
DB 167 GGGATAAATGAGC 149  
RESULT 9  
A1676097 446 bp mRNA EST 19-MAY-1999  
LOCUS A1676097  
DEFINITION wc05c12.x1 NCI-CGAP\_Prl28 Homo sapiens cDNA clone IMAGE:2314294 3'  
similar to gb:X06948 HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR  
ALPHA-SUBUNIT (HUMAN);, mRNA sequence.  
ACCESSION A1676097  
NID 94876577  
VERSION A1676097.1 GI:4876577  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 446)

**AUTHORS** NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
**JOURNAL** Unpublished (1997)  
**COMMENT** On Dec 20, 1995 this sequence version replaced gi:1133931.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bdrrp/image/image.html](http://www-bio.llnl.gov/bdrrp/image/image.html)

**FEATURES** Seq primer: -40UP from Gibco.  
**SOURCE** Location/Qualifiers  
1..446  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="1"  
/clone="IMAGE:2314294"  
/clone\_lib="NCI-CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI-CGAP\_Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (clonids  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

**BASE COUNT** 139 a 74 c 75 g 158 t

**Query Match** 75.2%; Score 15.8; DB 50; Length 446;  
**Best Local Similarity** 89.5%; Pred. No. 5.5e+02;  
**Matches** 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**QY** 1 tgggataataattggagc 19  
|||||  
**Db** 421 TGGGATTAATAATTGTAGC 439

**RESULT** 10  
**LOCUS** A1685796 540 bp mRNA EST 27-MAY-1999  
**DEFINITION** t420g08.x1 NCI-CGAP\_Pr28 Homo sapiens CDNA clone IMAGE:2251646 3'  
similar to gb:X06948 HIGH AFFINITY IMMUNOGLOBULIN EPSILON RECEPTOR  
ALPHA-SUBUNIT (HUMAN);, mRNA sequence.

**ACCESSION** A1685796  
**NID** 94897090  
**VERSION** A1685796.1 GI:4897090  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

**REFERENCE** 1 (bases 1 to 540)  
**AUTHORS** NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
**JOURNAL** Unpublished (1997)  
**COMMENT** On Feb 18, 1999 this sequence version replaced gi:4314767.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bdrrp/image/image.html](http://www-bio.llnl.gov/bdrrp/image/image.html)

**FEATURES** Trace considered overall poor quality  
**SOURCE** Seq primer: -40UP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers  
1..540  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2251646"  
/clone\_lib="NCI-CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI-CGAP\_Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (clonids  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

**BASE COUNT** 156 a 100 c 104 g 180 t

**Query Match** 75.2%; Score 15.8; DB 50; Length 540;  
**Best Local Similarity** 89.5%; Pred. No. 5.2e+02;  
**Matches** 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**QY** 1 tgggataataattggagc 19  
|||||  
**Db** 416 TGGGATTAATAATTGTAGC 434

**RESULT** 11  
**LOCUS** W92119 554 bp mRNA EST 02-FEB-1997  
**DEFINITION** zn48a09.r1 Soares\_fetal\_liver\_spleen\_INF5\_S1 Homo sapiens cDNA  
clone IMAGE:415288 5' similar to gb:X75755\_rna2 SPLICING FACTOR  
SC35 (HUMAN);, mRNA sequence.

**ACCESSION** W92119  
**NID** 61424502  
**VERSION** W92119.1 GI:1424502  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

**REFERENCE** 1 (bases 1 to 554)  
**AUTHORS** Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B.,  
Chisoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rinkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohlmann, P., Watson, R., Wilson, R.  
and Marra, M.  
**TITLE** Generation and analysis of 280,000 human expressed sequence tags  
**JOURNAL** Genome Res. 6 (9), 807-828 (1996)  
**MEDLINE** 97044478  
**COMMENT** On Oct 18, 1995 this sequence version replaced gi:1023832.

Contact: Wilson RK



Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewartson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 894 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 368.  
Location/Qualifiers

FEATURES  
source  
1. .554  
/organism="Homo sapiens"  
/db\_xref="GDB:1323758"  
/db\_xref="taxon:9606"  
/clone="IMAGE:415288"  
/clone\_id="Soares\_fetal\_liver\_spleen\_1NPLS.S1"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pTZ19 (Pharmacia)  
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI.  
This is a subtracted version of the original Soares fetal  
liver spleen INFLS library. 1st strand cDNA was primed  
with a Pac I - oligo(dT) primer [5',  
ACTGGAAGATTAATTAAGATCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pTZ19 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT  
ORIGIN  
157 a 103 c 199 t 5 others

Query Match 73.3% Score 15.4; DB 26; Length 554;  
Best Local Similarity 88.9% Pred. No. 7.9e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gataaataatcgagcgt 21  
|||||||

Db 165 GATAAATATGAGCAT 182

RESULT 12  
AA574291 421 bp mRNA EST 12-SEP-1997  
LOCUS n458a09.s1 NCI\_CGAP\_Pr2 Homo sapiens cDNA clone IMAGE:916696, mRNA  
DEFINITION  
sequence.  
ACCESSION AA574291  
NID 92348806  
VERSION AA574291.1 GI:2348806  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 421)  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Unpublished (1997)  
JOURNAL  
COMMENT On Apr 14, 1993 this sequence version replaced gi:693323.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nlh.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.lnl.gov/bdrp/image/image.html  
Insert Length: 1122 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 381.  
Location/Qualifiers

FEATURES  
source  
1. .421  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="19"  
/clone="IMAGE:916696"  
/clone\_id="NCI\_CGAP\_Pr2"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: PAMPI0; Site 1: NotI; Site 2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected preneoplastic cells  
histologically-determined to be prostatic interepithelial  
neoplasia 2 (PIN2) cells. Double-stranded cDNA was  
ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
cDNA with an adaptor-specific primer, and the resulting  
PCR product subcloned into PAMPI0 by the UDG-cloning  
method (Life Technologies). Average insert size is 600  
bp. NOTE: Not directionally cloned. This library was  
constructed by David Kitzman."

BASE COUNT  
ORIGIN  
133 a 72 c 65 g 151 t

Query Match 73.3% Score 15.4; DB 35; Length 421;  
Best Local Similarity 94.1% Pred. No. 8.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tgggataaataatcgga 17  
|||||||

Db 276 TGGATAAATATGAA 292

RESULT 13  
AA574359 375 bp mRNA EST 29-AUG-1997  
LOCUS n46a09.s1 NCI\_CGAP\_Pr2 Homo sapiens cDNA clone IMAGE:916792, mRNA  
DEFINITION  
sequence.  
ACCESSION AA574359  
NID 92348874  
VERSION AA574359.1 GI:2348874  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 375)  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Unpublished (1997)  
JOURNAL  
COMMENT On Apr 14, 1993 this sequence version replaced gi:693408.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nlh.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.lnl.gov/bdrp/image/image.html

Seq primer: -40m13 fwd. Et from Amersham  
High quality sequence stop: 341.  
Location/Qualifiers

1. 375

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="X"

/clone\_image="IMAGE:916792"

/clone\_lib="NCI\_CGAP\_P12"

/sex="Male"

/dev\_stage="45 years old"

/lab\_host="DH10B"

/note="vector: PAMP10; site\_1: NotI; site\_2: EcoRI; 1st

strand cDNA was primed with oligo(dT)17 on 50 ng of

DNase-treated, total cellular RNA obtained from

5,000-10,000 microdissected preneoplastic cells

histologically-determined to be prostatic intraepithelial

neoplasia 2 (PIN2) cells. Double-stranded cDNA was

ligated to EcoRI adaptors, 5 cycles of PCR applied to the

cDNA with an adaptor-specific primer, and the resulting

PCR product subcloned into PAMP10 by the UDG-cloning

method (Life Technologies). Average insert size is 600

bp. NOTE: Not directionally cloned. This library was

constructed by David Krizman."

BASE COUNT 120 a 65 c 58 g 132 t

ORIGIN

Query Match 73.3%; Score 15.4; DB 35; Length 375;

Best Local Similarity 94.1%; Pred. No. 8.8e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 tgggataaataatgga 17

|||||

Db 275 TGGGATTAATATTGAA 291

RESULT 14  
AI218562/c  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AI218562 438 bp mRNA EST 17-MAR-1999

qh19e01.x1 Soares\_NFL\_T\_GBC.S1 Homo sapiens cDNA clone

IMAGE:1845144 3', mRNA sequence.

AI218562

93798377

AI218562.1 GI:3798377

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 438)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On Jan 19, 1998 this sequence version replaced gi:2150572.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert length: 1128 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 422.

Location/Qualifiers

1. 438

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="1, 2q, 14; 1, 2q, 14; 21q"

/clone\_image="IMAGE:1845144"

/clone\_lib="Soares\_NFL\_T\_GBC.S1"

/lab\_host="DH10B"

BASE COUNT 150 a 69 c 63 g 134 t

ORIGIN

Query Match 73.3%; Score 15.4; DB 43; Length 438;

Best Local Similarity 94.1%; Pred. No. 8.5e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 tgggataaataatgga 17

|||||

Db 270 TGGGATTAATATTGAA 254

RESULT 15  
AI350781/c  
LOCUS  
DEFINITION  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AI350781 416 bp mRNA EST 30-DEC-1998

gg21d04.x1 Soares\_NHMPu\_S1 Homo sapiens cDNA clone IMAGE:1933159

3', mRNA sequence.

AI350781

94087987

AI350781.1 GI:4087987

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 416)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On Aug 21, 1998 this sequence version replaced.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Gibco

High quality sequence stop: 415.

Location/Qualifiers

1. 416

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="IMAGE:1933159"

/clone\_lib="Soares\_NHMPu\_S1"

/tissue\_type="Pooled human melanocyte, fetal heart, and

pregnant uterus"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBH19W, testis NHT, and B-cell  
NCI-CGAP-GCB1) were mixed, and ss circles were used in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive cDNA hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 157 a 73 c 68 g 140 t

ORIGIN

Query Match 73.3%; Score 15.4; DB 43; Length 438;

Best Local Similarity 94.1%; Pred. No. 8.5e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 tgggataaataatgga 17

|||||

Db 270 TGGGATTAATATTGAA 254

RESULT 15  
AI350781 416 bp mRNA EST 30-DEC-1998

gg21d04.x1 Soares\_NHMPu\_S1 Homo sapiens cDNA clone IMAGE:1933159

3', mRNA sequence.

AI350781

94087987

AI350781.1 GI:4087987

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 416)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On Aug 21, 1998 this sequence version replaced.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Gibco

High quality sequence stop: 415.

Location/Qualifiers

1. 416

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="IMAGE:1933159"

/clone\_lib="Soares\_NHMPu\_S1"

/tissue\_type="Pooled human melanocyte, fetal heart, and

pregnant uterus"

/lab\_host="DH10B"

/note="Organ: mixed (see below); Vector: p7T3D-Pac

(Pharmacia) with a modified polylinker; Site 1: Not I;

Site 2: Eco RI. Equal amounts of plasmid DNA from three

normalized libraries (melanocyte 2NBH, pregnant uterus

NBHPu, and fetal heart NBH19W) were mixed, and ss circles

were used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from pools of

5,000 clones made from the same 3 libraries. The pools

consisted of I.M.A.G.E. clones 26232-265223,

340488-345479, and 484468-489479."

BASE COUNT 150 a 69 c 63 g 134 t

ORIGIN

Query Match 73.3%; Score 15.4; DB 43; Length 438;

Best Local Similarity 94.1%; Pred. No. 8.5e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 tgggataaataatgga 17

|||||

Db 270 TGGGATTAATATTGAA 254

ORIGIN

Query Match 73.3%; Score 15.4; DB 45; Length 416;  
 Best Local Similarity 94.1%; Pred. No. 8.6e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 tgggataataattgga 17  
 ||||||||||||  
 Db 269 TCGGATAAATAATTGAA 253

Search completed: August 18, 1999, 17:04:50  
 Job time: 2796 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 17:17:40 ; Search time 543.6 Seconds  
(without alignments)  
122.860 Million cell updates/sec

Title: US-09-004-395-6

Perfect score: 21  
Sequence: 1 ctaatttcgagatgattc 21

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_v1:\*  
18: em\_fun:\*  
19: em\_hlg:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_v1:\*  
34: gb\_hcg1:\*  
35: gb\_hcg2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pr4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	21	100.0	332	1	BBCHEGEN	X91907 B.burgdorferi

c 2	21	100.0	13271	2	AE001168	AE001168 Borrelia
c 3	21	100.0	2700	2	BBU28962	U28962 Borrelia bu
c 4	21	100.0	1655	2	BBU62900	U62900 Borrelia bu
c 5	18.4	87.6	110000	34	CEY73P8_2	Continuation (3 of
c 6	17.8	84.8	1366	7	L107938	Y07938 L.inundata
c 7	17.8	84.8	3022	7	SDCAP2GEN	X94370 S.douglasii
c 8	17.8	84.8	1347	8	HHU87072	U87072 Haplophilic
c 9	17.8	84.8	156811	11	AC005548	AC005548 Homo sapi
c 10	17.4	82.9	86263	7	AB006702	AB006702 Arabidops
c 11	17.4	82.9	13166	7	AB025642	AB025642 Arabidops
c 12	17.4	82.9	1320	8	AF005536	AF005536 Orthocric
c 13	17.4	82.9	1320	8	AF005539	AF005539 Utiota obt
c 14	17.4	82.9	1320	8	AF005540	AF005540 Utiota lut
c 15	17.4	82.9	1320	8	AF005541	AF005541 Bryodioxon
c 16	17.4	82.9	90547	9	HS265314	Z84477 Human DNA s
c 17	17.4	82.9	36750	36	CELT1557	AF002985 Caenorhab
c 18	16.8	80.0	16947	2	AE000602	AE000602 Helicobac
c 19	16.8	80.0	2660	2	AF032910	AF032910 Leptospir
c 20	16.8	80.0	1495	7	D89119	D89119 Schistosach
c 21	16.8	80.0	93489	8	AC005359	AC005359 REVERSE-C
c 22	16.8	80.0	118737	8	T3H13	AF128396 Arabidops
c 23	16.8	80.0	100000	9	AP000021	AP000021 Homo sapi
c 24	16.8	80.0	100000	9	AP000163	AP000163 Homo sapi
c 25	16.8	80.0	146358	9	HS3D11	AL035088 Human DNA
c 26	16.8	80.0	34024	36	CEC18E9	Z70034 Caenorhabdi
c 27	16.8	80.0	36993	36	CELT10H9	AF067949 Caenorhab
c 28	16.4	78.1	1608	1	MEFTRTF	X70784 M.fervidus
c 29	16.4	78.1	20569	2	AE000636	AE000636 Helicobac
c 30	16.4	78.1	5061	7	NTNEDNA	X71020 N.tadacum N
c 31	16.4	78.1	113566	8	AC002330	AC002330 Arabidops
c 32	16.4	78.1	133191	8	AC007123	AC007123 Arabidops
c 33	16.4	78.1	107931	8	AC007478	AC007478 Arabidops
c 34	16.4	78.1	256172	35	AC005139	AC005139 Plasmodiu
c 35	16.4	78.1	210000	35	AC006839	AC006839 Homo sapi
c 36	16.4	78.1	196854	35	AC007684	AC007684 Homo sapi
c 37	16.4	78.1	1807	36	ECU47679	U47679 Euploies cr
c 38	16.4	78.1	4865	36	EUPWIC3	M28500 Euploies cr
c 39	16.2	77.1	1536	5	E05288	E05288 DNA encodin
c 40	16.2	77.1	1341	7	AARBCLGN	Z69221 A.africanus
c 41	16.2	77.1	10982	7	AB003364	AB003364 Adiantum
c 42	16.2	77.1	1341	7	ACRBL	Z69220 A.campanula
c 43	16.2	77.1	1341	7	AMZ77259	Z77259 A.monspelle
c 44	16.2	77.1	1624	7	BZCPRBCL	L11056 Bazania tr
c 45	16.2	77.1	14951	42	AC004970	AC004970 Homo sapi

## ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	MID	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL
BBCHEGEN/c	BBCHEGEN	B.burgdorferi chea gene.	X91907	91019753	X91907.1	GI:1019753	chea gene; chea protein; chemotactic response protein; histidine kinase.	Lyme disease spirochete.	Borrelia burgdorferi	Trueba,G.A., Old,I.G., Saint Girons,I. and Johnson,R.C.	98438936	Submitted (03-OCT-1995)	I.G. Old, Institut Pasteur, Bacteriology & Mycology, 28 rue du Dr Roux, 75724 Paris cedex 15, FRANCE	Old,I.G.	Direct Submission
BBCHEGEN/c	BBCHEGEN	B.burgdorferi chea gene.	X91907	91019753	X91907.1	GI:1019753	chea gene; chea protein; chemotactic response protein; histidine kinase.	Lyme disease spirochete.	Borrelia burgdorferi	Trueba,G.A., Old,I.G., Saint Girons,I. and Johnson,R.C.	98438936	Submitted (03-OCT-1995)	I.G. Old, Institut Pasteur, Bacteriology & Mycology, 28 rue du Dr Roux, 75724 Paris cedex 15, FRANCE	Old,I.G.	Direct Submission

FEATURES	source
1. 1332	location/Qualifiers
/organism="Borrelia burgdorferi"	
/strain="212"	
/db_xref="taxon:139"	
/clone_lib="Cos1"	
1. 202	
/gene="orf"	
<1. 202	
/gene="orf"	
/codon_start=2	
/transl_table=11	
/protein_id="CA63001.1"	
/db_xref="PID:g1019754"	
/db_xref="GI:1019754"	
/db_xref="SPTREMBL:O44876"	
/translation="DSDIDSESVFKYETSGTESLRKLAHETFKRVKLAKLEKISIAE	
GSQNKVEKTESKEPSESPKN"	
205. 256	
/note="hairpin"	
258. 261	
270. 332	
/gene="chea"	
270. >332	
/gene="chea"	
/function="chemotactic response"	
/codon_start=1	
/transl_table=11	
/product="histidine kinase"	
/protein_id="CA63002.1"	
/db_xref="PID:g1019755"	
/db_xref="GI:1019755"	
/db_xref="SWISS-PROT:Q44737"	
/translation="MEIDLLENBELGVEFEAQN"	
BASE COUNT	130 a 37 c 61 g 104 t
ORIGIN	
Query Match	100.0%; Score 21; DB 1; Length 332;
Best Local Similarity	100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ctaattttcgagatgattc 21
Db	202 CTAATTTCGAGATGATTTC 182
RESULT 2	
AE001168/c	13271 bp DNA BCT 15-DEC-1997
LOCUS	Borrelia burgdorferi (section 54 of 70) of the complete genome.
ACCESSION	AE001168 AE000783
NID	92686598
VERSION	AE001168.1 GI:2688598
SOURCE	Lyme disease spirochete.
ORGANISM	Borrelia burgdorferi
REFERENCE	Bacteria: Spirochaetales: Spirochaetaceae: Borrelia; Borrelia
AUTHORS	burgdorferi group.
	1 (bases 1 to 13271)
	Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
	Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
	Gatlin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
	Salzberg,S., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
	Gocayne,J.D., Weidman,J., Utterback,T., Wathey,L., McDonald,L.,
	Artlich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
	Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
	Genomic sequence of a Lyme disease spirochete, Borrelia
	burgdorferi
	Nature 390 (6660), 580-586 (1997)
JOURNAL	98065943
MEDLINE	2 (bases 1 to 13271)
REFERENCE	

FEATURES	AUTHORS
1. 1332	location/Qualifiers
/organism="Borrelia burgdorferi"	
/strain="212"	
/db_xref="taxon:139"	
/clone_lib="Cos1"	
1. 202	
/gene="orf"	
<1. 202	
/gene="orf"	
/codon_start=2	
/transl_table=11	
/protein_id="CA63001.1"	
/db_xref="PID:g1019754"	
/db_xref="GI:1019754"	
/db_xref="SPTREMBL:O44876"	
/translation="DSDIDSESVFKYETSGTESLRKLAHETFKRVKLAKLEKISIAE	
GSQNKVEKTESKEPSESPKN"	
205. 256	
/note="hairpin"	
258. 261	
270. 332	
/gene="chea"	
270. >332	
/gene="chea"	
/function="chemotactic response"	
/codon_start=1	
/transl_table=11	
/product="histidine kinase"	
/protein_id="CA63002.1"	
/db_xref="PID:g1019755"	
/db_xref="GI:1019755"	
/db_xref="SWISS-PROT:Q44737"	
/translation="MEIDLLENBELGVEFEAQN"	
BASE COUNT	130 a 37 c 61 g 104 t
ORIGIN	
Query Match	100.0%; Score 21; DB 1; Length 332;
Best Local Similarity	100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ctaattttcgagatgattc 21
Db	202 CTAATTTCGAGATGATTTC 182
RESULT 2	
AE001168/c	13271 bp DNA BCT 15-DEC-1997
LOCUS	Borrelia burgdorferi (section 54 of 70) of the complete genome.
ACCESSION	AE001168 AE000783
NID	92686598
VERSION	AE001168.1 GI:2688598
SOURCE	Lyme disease spirochete.
ORGANISM	Borrelia burgdorferi
REFERENCE	Bacteria: Spirochaetales: Spirochaetaceae: Borrelia; Borrelia
AUTHORS	burgdorferi group.
	1 (bases 1 to 13271)
	Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
	Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
	Gatlin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
	Salzberg,S., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
	Gocayne,J.D., Weidman,J., Utterback,T., Wathey,L., McDonald,L.,
	Artlich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
	Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
	Genomic sequence of a Lyme disease spirochete, Borrelia
	burgdorferi
	Nature 390 (6660), 580-586 (1997)
JOURNAL	98065943
MEDLINE	2 (bases 1 to 13271)
REFERENCE	

FEATURES	AUTHORS
1. 1332	location/Qualifiers
/organism="Borrelia burgdorferi"	
/strain="212"	
/db_xref="taxon:139"	
/clone_lib="Cos1"	
1. 202	
/gene="orf"	
<1. 202	
/gene="orf"	
/codon_start=2	
/transl_table=11	
/protein_id="CA63001.1"	
/db_xref="PID:g1019754"	
/db_xref="GI:1019754"	
/db_xref="SPTREMBL:O44876"	
/translation="DSDIDSESVFKYETSGTESLRKLAHETFKRVKLAKLEKISIAE	
GSQNKVEKTESKEPSESPKN"	
205. 256	
/note="hairpin"	
258. 261	
270. 332	
/gene="chea"	
270. >332	
/gene="chea"	
/function="chemotactic response"	
/codon_start=1	
/transl_table=11	
/product="histidine kinase"	
/protein_id="CA63002.1"	
/db_xref="PID:g1019755"	
/db_xref="GI:1019755"	
/db_xref="SWISS-PROT:Q44737"	
/translation="MEIDLLENBELGVEFEAQN"	
BASE COUNT	130 a 37 c 61 g 104 t
ORIGIN	
Query Match	100.0%; Score 21; DB 1; Length 332;
Best Local Similarity	100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ctaattttcgagatgattc 21
Db	202 CTAATTTCGAGATGATTTC 182
RESULT 2	
AE001168/c	13271 bp DNA BCT 15-DEC-1997
LOCUS	Borrelia burgdorferi (section 54 of 70) of the complete genome.
ACCESSION	AE001168 AE000783
NID	92686598
VERSION	AE001168.1 GI:2688598
SOURCE	Lyme disease spirochete.
ORGANISM	Borrelia burgdorferi
REFERENCE	Bacteria: Spirochaetales: Spirochaetaceae: Borrelia; Borrelia
AUTHORS	burgdorferi group.
	1 (bases 1 to 13271)
	Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
	Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
	Gatlin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
	Salzberg,S., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
	Gocayne,J.D., Weidman,J., Utterback,T., Wathey,L., McDonald,L.,
	Artlich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
	Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
	Genomic sequence of a Lyme disease spirochete, Borrelia
	burgdorferi
	Nature 390 (6660), 580-586 (1997)
JOURNAL	98065943
MEDLINE	2 (bases 1 to 13271)
REFERENCE	

FEATURES	AUTHORS
1. 1332	location/Qualifiers
/organism="Borrelia burgdorferi"	
/strain="212"	
/db_xref="taxon:139"	
/clone_lib="Cos1"	
1. 202	
/gene="orf"	
<1. 202	
/gene="orf"	
/codon_start=2	
/transl_table=11	
/protein_id="CA63001.1"	
/db_xref="PID:g1019754"	
/db_xref="GI:1019754"	
/db_xref="SPTREMBL:O44876"	
/translation="DSDIDSESVFKYETSGTESLRKLAHETFKRVKLAKLEKISIAE	
GSQNKVEKTESKEPSESPKN"	
205. 256	
/note="hairpin"	
258. 261	
270. 332	
/gene="chea"	
270. >332	
/gene="chea"	
/function="chemotactic response"	
/codon_start=1	
/transl_table=11	
/product="histidine kinase"	
/protein_id="CA63002.1"	
/db_xref="PID:g1019755"	
/db_xref="GI:1019755"	
/db_xref="SWISS-PROT:Q44737"	
/translation="MEIDLLENBELGVEFEAQN"	
BASE COUNT	130 a 37 c 61 g 104 t
ORIGIN	
Query Match	100.0%; Score 21; DB 1; Length 332;
Best Local Similarity	100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ctaattttcgagatgattc 21
Db	202 CTAATTTCGAGATGATTTC 182
RESULT 2	
AE001168/c	13271 bp DNA BCT 15-DEC-1997
LOCUS	Borrelia burgdorferi (section 54 of 70) of the complete genome.
ACCESSION	AE001168 AE000783
NID	92686598
VERSION	AE001168.1 GI:2688598
SOURCE	Lyme disease spirochete.
ORGANISM	Borrelia burgdorferi
REFERENCE	Bacteria: Spirochaetales: Spirochaetaceae: Borrelia; Borrelia
AUTHORS	burgdorferi group.
	1 (bases 1 to 13271)
	Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
	Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
	Gatlin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
	Salzberg,S., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
	Gocayne,J.D., Weidman,J., Utterback,T., Wathey,L., McDonald,L.,
	Artlich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
	Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
	Genomic sequence of a Lyme disease spirochete, Borrelia
	burgdorferi
	Nature 390 (6660), 580-586 (1997)
JOURNAL	98065943
MEDLINE	2 (bases 1 to 13271)
REFERENCE	

FEATURES	AUTHORS
1. 1332	location/Qualifiers
/organism="Borrelia burgdorferi"	
/strain="212"	
/db_xref="taxon:139"	
/clone_lib="Cos1"	
1. 202	
/gene="orf"	
<1. 202	
/gene="orf"	
/codon_start=2	
/transl_table=11	
/protein_id="CA63001.1"	
/db_xref="PID:g1019754"	
/db_xref="GI:1019754"	
/db_xref="SPTREMBL:O44876"	
/translation="DSDIDSESVFKYETSGTESLRKLAHETFKRVKLAKLEKISIAE	
GSQNKVEKTESKEPSESPKN"	
205. 256	
/note="hairpin"	
258. 261	
270. 332	
/gene="chea"	
270. >332	
/gene="chea"	
/function="chemotactic response"	
/codon_start=1	
/transl_table=11	
/product="histidine kinase"	
/protein_id="CA63002.1"	
/db_xref="PID:g1019755"	
/db_xref="GI:1019755"	
/db_xref="SWISS-PROT:Q44737"	
/translation="MEIDLLENBELGVEFEAQN"	
BASE COUNT	130 a 37 c 61 g 104 t
ORIGIN	
Query Match	100.0%; Score 21; DB 1; Length 332;
Best Local Similarity	100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ctaattttcgagatgattc 21
Db	202 CTAATTTCGAGATGATTTC 182
RESULT 2	
AE001168/c	13271 bp DNA BCT 15-DEC-1997
LOCUS	Borrelia burgdorferi (section 54 of 70) of the complete genome.
ACCESSION	AE001168 AE000783
NID	92686598
VERSION	AE001168.1 GI:2688598
SOURCE	Lyme disease spirochete.
ORGANISM	Borrelia burgdorferi
REFERENCE	Bacteria: Spirochaetales: Spirochaetaceae: Borrelia; Borrelia
AUTHORS	burgdorferi group.
	1 (bases 1 to 13271)
	Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
	Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
	Gatlin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
	Salzberg,S., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
	Gocayne,J.D., Weidman,J., Utterback,T., Wathey,L., McDonald,L.,
	Artlich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
	Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
	Genomic sequence of a Lyme disease spirochete, Borrelia
	burgdorferi
	Nature 390 (6660), 580-586 (1997)
JOURNAL	98065943
MEDLINE	2 (bases 1 to 13271)
REFERENCE	

FEATURES	AUTHORS
1. 1332	location/Qualifiers
/organism="Borrelia burgdorferi"	
/strain="212"	
/db_xref="taxon:139"	
/clone_lib="Cos1"	
1. 202	
/gene="orf"	
<1. 202	
/gene="orf"	
/codon_start=2	
/transl_table=11	
/protein_id="CA63001.1"	
/db_xref="PID:g1019754"	
/db_xref="GI:1019754"	
/db_xref="SPTREMBL:O44876"	
/translation="DSDIDSESVFKYETSGTESLRKLAHETFKRVKLAKLEKISIAE	
GSQNKVEKTESKEPSESPKN"	
205. 256	
/note="hairpin"	
258. 261	
270. 332	
/gene="chea"	
270. >332	
/gene="chea"	
/function="chemotactic response"	
/codon_start=1	
/transl_table=11	
/product="histidine kinase"	
/protein_id="CA63002.1"	
/db_xref="PID:g1019755"	
/db_xref="GI:1019755"	
/db_xref="SWISS-PROT:Q44737"	
/translation="MEIDLLENBELGVEFEAQN"	
BASE COUNT	130 a 37 c 61 g 104 t
ORIGIN	
Query Match	100.0%; Score 21; DB 1; Length 332;
Best Local Similarity	100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ctaattttcgagatgattc 21
Db	202 CTAATTTCGAGATGATTTC 182
RESULT 2	
AE001168/c	13271 bp DNA BCT 15-DEC-1997
LOCUS	Borrelia burgdorferi (section 54 of 70) of the complete genome.
ACCESSION	AE001168 AE000783
NID	92686598
VERSION	AE001168.1 GI:2688598
SOURCE	Lyme disease spirochete.
ORGANISM	Borrelia burgdorferi
REFERENCE	Bacteria: Spirochaetales: Spirochaetaceae: Borrelia; Borrelia
AUTHORS	burgdorferi group.
	1 (bases 1 to 13271)
	Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
	Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
	Gatlin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
	Salzberg,S., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
	Gocayne,J.D., Weidman,J., Utterback,T., Wathey,L., McDonald,L.,
	Artlich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
	Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
	Genomic sequence of a Lyme disease spirochete, Borrelia
	burgdorferi
	Nature 390 (6660), 580-586 (1997)
JOURNAL	98065943
MEDLINE	2 (bases 1 to 13271)
REFERENCE	

FEATURES	AUTHORS
1. 1332	location/Qualifiers
/organism="Borrelia burgdorferi"	
/strain="212"	
/db_xref="taxon:139"	
/clone_lib="Cos1"	
1. 202	
/gene="orf"	
<1. 202	
/gene="orf"	
/codon_start=2	
/transl_table=11	
/protein_id="CA63001.1"	
/db_xref="PID:g1019754"	
/db_xref="GI:1019754"	
/db_xref="SPTREMBL:O44876"	
/translation="DSDIDSESVFKYETSGTESLRKLAHETFKRVKLAKLEKISIAE	
GSQNKVEKTESKEPSESPKN"	
205. 256	
/note="hairpin"	
258. 261	
270. 332	
/gene="chea"	
270. >332	
/gene="chea"	
/function="chemotactic response"	
/codon_start=1	
/transl_table=11	
/product="histidine kinase"	
/protein_id="CA63002.1"	
/db_xref="PID:g1019755"	
/db_xref="GI:1019755"	
/db_xref="SWISS-PROT:Q44737"	
/translation="MEIDLLENBELGVEFEAQN"	
BASE COUNT	130 a 37 c 61 g 104 t
ORIGIN	
Query Match	100.0%; Score 21; DB 1; Length 332;
Best Local Similarity	100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ctaattttcgagatgattc 21
Db	202 CTAATTTCGAGATGATTTC 182
RESULT 2	
AE001168/c	13271 bp DNA BCT 15-DEC-1997
LOCUS	Borrelia burgdorferi (section 54 of 70) of the complete genome.
ACCESSION	AE001168 AE000783
NID	92686598
VERSION	AE001168.1 GI:2688598
SOURCE	Lyme disease spirochete.
ORGANISM	Borrelia burgdorferi
REFERENCE	Bacteria: Spirochaetales: Spirochaetaceae: Borrelia; Borrelia
AUTHORS	burgdorferi group.
	1 (bases 1 to 13271)
	Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
	Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
	Gatlin,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
	Salzberg,S., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
	Gocayne,J.D., Weidman,J., Utterback,T., Wathey,L., McDonald,L.,
	Artlich,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
	Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
	Genomic sequence of a Lyme disease spirochete, Borrelia
	burgdorferi
	Nature 390 (6660), 580-586 (1997)
JOURNAL	98065943
MEDLINE	2 (bases 1 to 13271)
REFERENCE	

FEATURES	AUTHORS
1. 1332	location/Qualifiers
/organism="Borrelia burgdorferi"	
/strain="212"	
/db_xref="taxon:139"	
/clone_lib="Cos1"	
1. 202	
/gene="orf"	
<1. 202	
/gene="orf"	
/codon_start=2	
/transl_table=11	
/protein_id="CA63001.1"	
/db_xref="PID:g1019754"	
/db_xref="GI:1019754"	
/db_xref="SPTREMBL:O44876"	
/translation="DSDIDSESVFKYETSGTESLRKLAHETFKRVKLAKLEKISIAE	
GSQNKVEKTESKEPSESPKN"	
205. 256	
/note="hairpin"	
258. 261	
270. 332	
/gene="chea"	
270. >332	
/gene="chea"	
/function="chemotactic response"	
/codon_start=1	
/transl_table=11	
/product="histidine kinase"	
/protein_id="CA63002.1"	
/db_xref="PID:g1019755"	
/db_xref="GI:1019755"	
/db_xref="SWISS-PROT:Q44737"	
/translation="MEIDLLENBELGVEFEAQN"	
BASE COUNT	130 a 37 c 61 g 104 t
ORIGIN	
Query Match	100.0%; Score 21; DB 1; Length 332;

```
/db_xref="PID:g2688606"
/db_xref="GI:2688606"
/translacion-"MJOKEIYFGPKIIDDKNNSKLNPFKVSFELGSDHYLVLM
QVEIRKSSNFTYVPNAKRYVAGLDNIGETIIPIDIRIMENLEFNKJEDIVLKN
EDLLIGYIVDKINNVFSDISLDDPPHVSODSLIVIGVYDNEKLYILNVFPI
FNYGEEKLKPQNFYEKSDPVSODCNLDIEKSDKUSKSDGNENETLKN
TAAPNENIKKNLKIYFNSLVNDVLEKYGVAFNVDYDLPDSTLDEFYSKSG
NLWGADCLEEFKNEIVSRSEFNKNSIFNVEIGGSGKEMALANALSEYVKEP
KLAIDNDLSKVETSRVSESEIGISEYSRNSPQVYFKSEIISLNVFEYS
DALFSDLPDNLGWFELKDVLCFLDSKQIILNIIASKAIKALLVIGDNELKNNDV
FIKESKRYFNLYKKV"
5181..5666
/gene="BB0671"
5181..5666
/gene="BB0671"
/notes="similar to GP:1408273 percent identity: 98.76;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="chemotaxis operon protein (chex)"
/protein_id="AAC67022.1"
/db_xref="PID:g2688605"
/db_xref="GI:2688605"
/translacion-"MRDYEIPEFLDPAASVLRDMLVNTENMKRPGKLSINOKIGVS
VIVGLAGVSGSIIDMDIEFALFVASKLNFEEDDDEETKEMVATLVEGNIIA
GNFVTLHAKGFEDIPPAFYGENNKNKSGSEALIVPSPDGKIIIEVNAIRER
V"
5706..6146
/gene="BB0672"
5706..6146
/gene="BB0672"
/notes="similar to GP:1408274 percent identity: 97.95;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="chemotaxis response regulator (cher-3)"
/protein_id="AAC67021.1"
/db_xref="PID:g2688604"
/db_xref="GI:2688604"
/translacion-"MIQKTIADSSSKPRGINDYDGPENVLIVDSVFTVKOLTOI
FTSEGNIIIDTAADGEAVIKRKNHYNIDIVTIDIMPKMDGTCSNIMEPKNAR
VIMISALGRQIYVDCILKAKRTIVPLBRAKVLQRMVSIFYA"
complement(6187..6702)
/gene="BB0673"
complement(6187..6702)
/gene="BB0673"
/notes="similar to GP:1408275 percent identity: 97.56;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAC67020.1"
/db_xref="PID:g2688603"
/db_xref="GI:2688603"
/translacion-"MKIQIILMLALLDFPLNARLDISIEKRADEIRKYSNYLIL
EKEYVTNPSEIKENIKYKLEHFKVKSIMLNKYSLNSNYKANKYLLQSEIDIK
ELKXIKFKIKINICIFKSHSLIYKKGIFYKLELYIENNAEPLKIFNINITYEKLNDK
ISNEMIFFPRE"
complement(6699..7745)
/gene="BB0674"
complement(6699..7745)
/gene="BB0674"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=1
/product="B. burgdorferi predicted coding region BB0674"
/protein_id="AAC67027.1"
/db_xref="PID:g2688610"
/db_xref="GI:2688610"
/translacion-"MVKKFSIFLKAIIIFSEFELLIELSLIFLPYKIRALFLGF
LFTDIFIFILYIKAKYLSORLEIVYNNLFFDIIICLIPLAYSSYQLNNIIAAE
TILNPMLSLFLKRLRLRNLIIIEIYNSKKNLILAFARTFSMLIPTFPFI
```

```
ISSKRWNSIPKQEPNITKNISINERAYIKKYPFIILIEKODIYSKDEIFV
YSSPEYRVIEMETKIFYIDKYLQKRSDSLIGFLFLFASFTIFLMNFKKASFL
NPIILMTKLIKQDLEVRKIDIPFLSEEEKYELAKSFNNLLKREKLNKSKRSKIPLEI
EKRKIKNKQOEIK"
complement(7739..8602)
/gene="BB0675"
complement(7739..8602)
/gene="BB0675"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=1
/product="B. burgdorferi predicted coding region BB0675"
/protein_id="AAC67026.1"
/db_xref="PID:g2688609"
/db_xref="GI:2688609"
/translacion-"NMTRKPEFPLVYHFNHESIKSISALEKELEILNYKKNKITA
TFIKNDPESIKDIOVKDQDIMTIPVLSIGIEADENIVKPLSKELTKNDLNI
FNFVKVNSSLRKEFFYNNTISNGYITFVINKLEGNSTYIYLIOENKALYSSDI
KNYIKILLKLVIVIKYCEKGIETTKNIESTSKAISNDTDLDEKTAIILIESFFK
YETIQKSPISITLAIIFSARARTPKYKNNPVKGFIGYDESMFSIKSGSREYDSRIK
ELSEIAKVNKW"
complement(8599..9261)
/gene="BB0676"
complement(8599..9261)
/gene="BB0676"
/notes="similar to GB:1442023 SP:P44755 PID:1004013
PID:1222502 PID:1204815 percent identity: 25.58;
identified by sequence similarity; putative"
/codon_start=1

Query Match 100.0%; Score 21; DB 2; Length 13271;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctaatttcggagatgattc 21
|||||
Db 1100 CTAATTTTCGAGATGATTC 1080

RESULT 3
BBU28962/c BBU28962 2700 bp DNA BCT 16-JAN-1998
LOCUS Borrelia burgdorferi histidine kinase (chea) gene, complete cds.
DEFINITION U28962
ACCESSION G1113814
NID U28962.1 GI:1113814
VERSION 1
KEYWORDS
SOURCE
ORGANISM
Lyme disease spirochete.
Borrelia burgdorferi
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
burgdorferi group
1 (bases 1 to 2700)
Trueba,G.A., Old,I.G., Saint Girons,I. and Johnson,R.C.
A chea chew operon in Borrelia burgdorferi, the agent of Lyme
disease
Res. Microbiol. 148 (3), 191-200 (1997)
JOURNAL
MEDLINE 98438936
REFERENCE 2 (bases 1 to 2700)
AUTHORS Trueba,G.A. and Johnson,R.C.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1995) Gabriel A. Trueba, Microbiology, University
of Minnesota, 420 Delaware Street S.E., Minneapolis, MN 55455, USA
On Dec 12, 1995 this sequence version replaced gi:984522.
COMMENT
FEATURES
Location/Qualifiers
1..2700
/organism="Borrelia burgdorferi"
/strain="CT-1"
/db_xref="taxon:139"
95..2689
/gene="chea"
95..2689
/gene="chea"
/gene="chea"
```

```
/codon_start=1
/translation="involved in chemotaxis"
/function="histidine kinase"
/product="histidine kinase"
/protein_id="AAB96835.1"
/db_xref="PID:g1113815"
/db_xref="GI:1113815"
/translation="MEIIDLNEELIGVFEEAQNVLIDLEENIMSLIEDDNNSTID
EIFRAHTLKGSSASLDMELSDPFIHVEDFADIRDKVNIINDVLDSLIVK
EMALRIDGKYLDISDLKSKLKOFLVIDQTFKREDGNSIKNNFCLSDDEIR
EGDIGKVRISVFNSSNHEVNGGLKLENIKNGSVLHTIPKREOLIEKFPK
RDVYLISYSDIEGVKSLDSSNLIESTYDEENYKEELKRLADEIRIKDSDSNFVN
DNDFDTDEIDLDLEVENOKLFRDLFVNDPNAITSGLOALKSLKTRKSD
DSSELAKRFDVYLYLISNTSESIKRLNLPVVSHEPIKNNLESLSVRLED
DEAPFENKRNKNSPISVNLIRIDSKRIDLYLNVSEAVISKSSYNQINSEMTLFY
NENFYDQESFORNFIIDKIVFDAGLTLEDIESHINSLSMFKKELDKIDSELR
NSFFELONFMTSGRLSRITLDHESVLTBMLPISNIFSRFRVYVDLPKIKLTV
NLKMEGETEIDKVIDLVDPLMHCYNSMDHGEETVEEYKRGSKSAQIILIRAK
EGNVTIETEDGIGIDPKVIRKRIIEKGTIKEDTISDFELINLIRNPGSTAVQT
DLSGGVGLDVVKKSIKRLNGTILVSEIGLTFKIRKLPPLTVIIOGLAVKSGSEY
VPLNNVLETRITEHDIKLENYHEVNLNDEVYSLRLPLNITRDEDLIERFLI
VNTSNMKIALIVDSILGEEDFVYKPIKRFSSSAGIYGATTLGNGKVVLIIDVEKLF
DLOKDTKE"
```

BASE COUNT 988 a 278 c 507 g 927 t

ORIGIN

Query Match 100.0%; Score 21; DB 2; Length 2700;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctaatttcggagatgattc 21  
|||||  
Db 27 CTAATTTTCGAGATGATTC 7

RESULT 4  
LOCUS BB062900 1655 bp DNA BCT 15-JAN-1997  
DEFINITION Borrelia burgdorferi flagellar filament outsheath protein (flaA)  
gene, complete cds, and chemotaxis histidine kinase (cheA) gene,  
partial cds.  
ACCESSION U62900  
NID 91575445  
VERSION U62900.1 GI:1575445  
KEYWORDS  
SOURCE  
ORGANISM  
Lyme disease spirochete.  
Borrelia burgdorferi  
Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia  
burgdorferi group.  
1 (bases 1 to 1655)  
Ge,Y. and Charon,N.W.  
An unexpected flaA homolog is present and expressed in Borrelia  
burgdorferi.  
JOURNAL J. Bacteriol. 179 (2), 552-556 (1997)  
MEDLINE 97144545  
REFERENCE 2 (bases 1 to 1655)  
AUTHORS Ge,Y.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-1996) Yigong Ge, West Virginia University,  
Microbiology, HSCN, Morgantown, WV 26506, USA  
FEATURES  
source  
1..1655  
/organism="Borrelia burgdorferi"  
/strain="212"  
/db\_xref="taxon:139"  
/clone="pm1 and pm2"  
<1..395  
/function="unknown"  
/note="orfa"  
/codon\_start=3  
/transl\_table=11  
/protein\_id="AAC44769.1"

CDS

```
/db_xref="PID:g1575446"  
/db_xref="GI:1575446"  
/translation="DNLFSKRVFFHSSKGFVANLYRLRDEQNKDNLVLVDFLIG  
SNGESEGFILSDSRFLYSLKNGVYVNLRSREFYDSFNNGDYNESPDVYVNFAMSL  
IKMRFNYPGRIKRIIILVGCILKEGS"  
473..1498  
/gene="flaA"  
473..1498  
/gene="flaA"  
/function="putative flagellar filament outsheath protein"  
/translation="MEIIDLNEELIGVFEEAQNVLIDLEENI"  
1566..1655  
/gene="cheA"  
1566..>1655  
/gene="cheA"  
/function="chemotaxis histidine kinase"  
/codon_start=1  
/transl_table=11  
/product="FlaA protein"  
/protein_id="AAC44770.1"  
/db_xref="PID:g1575447"  
/db_xref="GI:1575447"  
/translation="MKRKASLIIFLLSTVLFAOETDGLAEGSKRAEPGLVDPAL  
ARBPSSRLDLTVYVYVYGAGCYKPEPMYVDLGINNSVLTLSBARQAVKNSV  
VAPVAVSESKRTAGDTILGVRVLFPSYSSAMIMPEKIPYSGSGNQIFGKGLI  
DNKTKMEIVSVYSLGELYDLVLEFEDNMGMEYASMGTLKFGWADLIWSPNIP  
NISRLIKDVPNPYPLASSKMRKAFVSKSSKYNFIYVYKDLRYLDKLSVSD  
SDIDSEGFVYETSGTESLRKLKAHETFRVLKLRKISIAEGSFQNFVEKIESEK  
EESSPKN"  
1566..1655  
/gene="cheA"  
1566..>1655  
/gene="cheA"  
/function="chemotaxis histidine kinase"  
/codon_start=1  
/transl_table=11  
/product="CheA protein"  
/protein_id="AAC44771.1"  
/db_xref="PID:g1575448"  
/db_xref="GI:1575448"  
/translation="MEIIDLNEELIGVFEEAQNVLIDLEENI"
```

BASE COUNT 560 a 182 c 308 g 605 t

ORIGIN

Query Match 100.0%; Score 21; DB 2; Length 1655;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctaatttcggagatgattc 21  
|||||  
Db 1498 CTAATTTTCGAGATGATTC 1478

RESULT 5  
CEY73F8\_2  
WPCOMMENT  
Sequence split into 4 fragments LOCUS CEY73F8 Accession AL022285  
Fragment Name Begin End  
CEY73F8\_0 1 110000  
CEY73F8\_1 100001 210000  
CEY73F8\_2 200001 310000  
CEY73F8\_3 300001 376956  
Continuation (3 of 4) of CEY73F8 from base 200001 (AL022285 Caenorhabditis elegans ch

Query Match 87.6%; Score 18.4; DB 34; Length 110000;  
Best Local Similarity 95.0%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 taatttcggagatgattc 21  
|||||  
Db 62561 TAATTTTCGAGATGATTC 62580

RESULT 6  
LOCUS LI07938 1366 bp DNA PLN 22-JAN-1998  
DEFINITION L. inuundata chloroplast rbcL gene.



[illegible]

REFERENCE	1 (bases 1 to 3022)
AUTHORS	Li, G. Y., Tian, G. L., Stonitski, P. P. and Herbert, C. J.
TITLE	The CBP2 gene from <i>Saccharomyces douglasii</i> is a functional homolog of the <i>Saccharomyces cerevisiae</i> gene and is essential for respiratory growth in the presence of a wild-type (ntron-containing) mitochondrial genome
JOURNAL	Mol. Gen. Genet. 250 (3), 316-322 (1996)
MEDLINE	96180648
REFERENCE	2 (bases 1 to 3022)
AUTHORS	Herbert, C. J.
TITLE	Direct Submission
JOURNAL	Submitted (20-DEC-1995) C. J. Herbert, Centre de Genetique Moleculaire du CNRS, F-91198 Gif-sur-Yvette, FRANCE
FEATURES	Location/Qualifiers
source	1..3022
gene	/organism="Saccharomyces douglasii"
	/db_xref="taxon:46617"
	775..2667
	/gene="CBP2"
CDS	775..2667
	/gene="CBP2"
	/codon_start=1
	/protein_id="CA64148.1"
	/db_xref="PID:e217090"
	/db_xref="PID:g1149548"
	/db_xref="GI:1149548"
	/db_xref="SPTREMBL:008996"
	/translation="MVMQTLFWYSLRGQSSSHRYRYKFMNENITHOVPRCKOAFKKANISYECDEGLKLYNASLMDLQKMLRDNAPDHYKIVRTLVKSSKRSIOHMEQIPIPHPLSLYENLLSEMDNPKPFEENQKLPFEGNDQKLGQVWPKRSISFNPDGICGYKMYNDRDLKONGNEVPMFVPAKRLPILEVGDADIGSPDGCIPYVVKRAAFVTVEDKLRITISPLCLNKKRTDKRANRRLLENKGEFFYDAKGAASERAAAGNAVTLKOLLERSVSHKTLWSKQNKRCQGLLRATILSNDPSINQLRTEFCNFIILNFTILQRRNKSIRDFSDNNVPSRFNMNWDYSIMKOYQAESMTLPADQASKLINRYKQYDFLDDLOTYSILVSEKMQSFQNNETSLSFEHTILQITLTKRSKIRIKPQMLYKFMODDLRATIMELTGFESINATLGPGRANOSLSANAKLKLANQLHAFBEKITYAEKFRVNRPIQLARPLTSTFKIYIILDKNAIPIPTFLKFTQITTTTPKDSSEVLEHGHMCIDKMDKSTFMILYELKTEDNEKAVNPQREKIVDNIIGLLSNDSEH"
BASE COUNT	986 a 575 c 595 g 866 t
ORIGIN	
Query Match	84.88; Score 17.8; DB 7; Length 3022;
Best Local Similarity	90.58; Pred. No. 73;
Matches	19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 ctaattttcgagatgattc 21
DB	
	2901 CTAATTATCGAGAGATTC 2881
RESULT	8
LOCUS	HHU87072
DEFINITION	HHU87072 1347 bp DNA PLN 28-JAN-1998
ACCESSION	Haplomitrium hookeri 1.5-bisphosphate
VERSION	carboxylase/oxygenase large subunit (rbcL) gene, chloroplast gene
KEYWORDS	encoding chloroplast protein, partial cds.
SOURCE	U87072
ORGANISM	62149494
	U87072.1 GI:2149494
REFERENCE	Haplomitrium hookeri.
AUTHORS	Chloroplast Haplomitrium hookeri
TITLE	Eukaryote: Viridiplantae; Streptophyta; Embryophyta;
	Marchantiophyta; Charophytales; Haplomitriaceae; Haplomitrium.
	1 (bases 1 to 1347)
	Lewis, L.A., Mishler, B.D. and Vilgalys, R.
	Phylogenetic relationships of the liverworts (Hepaticae), a basal
	embryophyte lineage, inferred from nucleotide sequence data of the
	chloroplast gene, rbcL
	Mol. Phylogenet. Evol. (1997) In press
JOURNAL	2 (bases 1 to 1347)



```

repeat_region /rpt_family="MER39b"
complement(17579..17813)
repeat_region /rpt_family="MLTIG"
18381..18915
repeat_region /rpt_family="NALA"
18917..19024
repeat_region /rpt_family="purine-rich"
20267..20318
repeat_region /rpt_family="LTR30"
20322..20856
repeat_region /rpt_family="LTR12"
21113..21475
repeat_region /rpt_family="THE1B"
complement(21829..21848)
repeat_region /rpt_family="(CAAA)n"
21928..22223
repeat_region /rpt_family="AluY"
22893..22932
repeat_region /rpt_family="POLY_A"
23451..23993
repeat_region /rpt_family="MER50"
complement(24128..24243)
repeat_region /rpt_family="L2"
complement(24353..24376)
repeat_region /rpt_family="POLY_A"
25078..25424
repeat_region /rpt_family="MLTIC"
complement(25425..25637)
repeat_region /rpt_family="AluSx"
complement(25879..25981)
repeat_region /rpt_family="AluSx"
25982..26087
repeat_region /rpt_family="MLTIC"
26096..26381
repeat_region /rpt_family="L2"
complement(26576..26847)
repeat_region /rpt_family="MLTIG"
26939..27037
repeat_region /rpt_family="MER5A"
27219..27278
repeat_region /rpt_family="MT-rich"
27491..27797
repeat_region /rpt_family="AluY"
28291..28550
repeat_region /rpt_family="AluSx"
complement(30350..30915)
repeat_region /rpt_family="MER41A"
31685..31789
repeat_region /rpt_family="L2"
31790..31830
repeat_region /rpt_family="(TGAA)n"
complement(31935..32642)
repeat_region /rpt_family="LTR8"
complement(36273..36584)
repeat_region /rpt_family="AluSg"
36594..36661
repeat_region /rpt_family="MLT1Y"
36820..36926
repeat_region /rpt_family="L2"
complement(36958..37175)
repeat_region /rpt_family="L1MA10"
complement(37176..37484)
repeat_region /rpt_family="AluY"
complement(37488..37617)
repeat_region /rpt_family="FLAM_A"
complement(37618..37919)
repeat_region /rpt_family="L1MA10"
complement(37920..38210)
repeat_region /rpt_family="AluSc"
complement(38211..38399)
repeat_region /rpt_family="L1MA10"
38400..38685
/rpt_family="AluSg"

repeat_region 38686..38719
/rpt_family="(TAAA)n"
complement(38721..38744)
repeat_region /rpt_family="L1MA10"
complement(38759..38931)
repeat_region /rpt_family="L1MC/D"
complement(39363..39654)
repeat_region /rpt_family="AluSx"
39890..40196
repeat_region /rpt_family="AluY"
complement(40072..40148)
/note="Single-stranded coverage."
repeat_region /rpt_family="L1ME3A"
complement(40433..40722)
repeat_region /rpt_family="L1ME3A"
complement(40724..40747)
repeat_region /rpt_family="(CAA)n"
complement(40748..41025)
repeat_region /rpt_family="AluY"
complement(41026..41492)
/rpt_family="L1ME3A"
complement(41141..41570)
/note="Single-stranded coverage."
repeat_region complement(41529..41697)
/rpt_family="MIR"

Query Match 84.8%; Score 17.8; DB 11; Length 156811;
Best Local Similarity 90.5%; Pred. No. 60;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaatttcgagatgattc 21
Db 41966 CTAATTTCTGAGATGACTC 41946

RESULT 10
AB006702 86263 bp DNA PLN 05-FEB-1999
LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, pl clone: MP012,
DEFINITION complete sequence.
ACCESSION AB006702
VERSION 92351067
NID AB006702.1 GI:2351067
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui pl
clone:MP012.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
REFERENCE 1 (sites)
AUTHORS Kotani,H., Nakamura,Y., Seto,S., Kaneko,T., Asamizu,E., Miyaajima,N.
and Tabata,S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. II.
Sequence features of the regions of 1,044,062 bp covered by
thirteen physically assigned pl clones
JOURNAL DNA Res. 4 (4), 291-300 (1997)
MEDLINE 98069011
REFERENCE 2 (bases 1 to 86263)
AUTHORS Nakamura,Y.
TITLE Direct Submision
JOURNAL Submitted (22-AUG-1997) to the DDBJ/EMBL/Genbank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2, 1532-3, Yana, Kisarazu, Chiba 292, Japan
(E-mail:ynakamukazusa.or.jp, Tel:+81-438-52-3935,
Fax:+81-438-52-3934)
FEATURES
source 1..86263
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MP012"

```

BASE COUNT 26950 a 15052 c 14436 g 29825 t  
 ORIGIN

Query Match 82.9%; Score 17.4; DB 7; Length 86263;  
 Best Local Similarity 94.7%; Pred. No. 96;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ctaatttcggagatgat 19  
 ||||||||||||||||  
 Db 4828 CTATTTTCGAGATGAT 4846

RESULT 11  
 AB025642  
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, pl clone: T32G24,  
 DEFINITION complete sequence.  
 ACCESSION AB025642  
 NID 94589451  
 VERSION AB025642.1 GI:4589451  
 KEYWORDS HTG.  
 SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone\_Mitsui PI

ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
 Arabidopsis.  
 1 (sites)  
 AUTHORS Nakamura,Y.  
 TITLE Structural Analysis of Arabidopsis thaliana Chromosome 5. XIII  
 JOURNAL Unpublished (1999)  
 REFERENCE 2 (bases 1 to 13186)  
 AUTHORS Nakamura,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-APR-1999) to the DDBJ/EMBL/GenBank databases.  
 Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of  
 Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 297, Japan  
 (E-mail:yinakamura@kazusa.or.jp, Tel:+81-438-52-3935,  
 Fax:+81-438-52-3934)

FEATURES  
 source  
 1..13186  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="5"  
 /clone="T32G24"  
 /clone\_lib="Mitsui PI"  
 BASE COUNT 4238 a 2304 c 2279 g 4365 t  
 ORIGIN

Query Match 82.9%; Score 17.4; DB 7; Length 13186;  
 Best Local Similarity 94.7%; Pred. No. 1.1e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 taatttcggagatgat 20  
 ||||||||||||||||  
 Db 12522 TAATTTTCGAGATGAT 12540

RESULT 12  
 AF005536  
 LOCUS AF005536 1320 bp DNA PLN 17-DEC-1998  
 DEFINITION Orthotrichum lyelli ribulose-bisphosphate carboxylase large subunit  
 (rbcL) gene, chloroplast gene encoding chloroplast protein, partial  
 cds.  
 ACCESSION AF005536  
 NID 94009399  
 VERSION AF005536.1 GI:4009399  
 KEYWORDS

SOURCE Orthotrichum lyelli.  
 ORGANISM Chloroplast Orthotrichum lyelli  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryopsida;  
 Bryidae; Orthotrichales; Orthotrichaceae; Orthotrichum.  
 REFERENCE  
 AUTHORS Goffinet,B., Bayer,R.J. and Vitt,D.H.  
 TITLE 1 (bases 1 to 1320)  
 JOURNAL Goffinet,B., Bayer,R.J. and Vitt,D.H.  
 REFERENCE Circumscription and phylogeny of the orthotrichales (Bryopsida)  
 Am. J. Bot. 85, 1324-1337 (1998)  
 2 (bases 1 to 1320)  
 AUTHORS Goffinet,B.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAY-1997) Department of Botany, Duke University,  
 P.O.Box 90339, Durham, NC 27708

FEATURES  
 source  
 1..1320  
 /organism="Orthotrichum lyelli"  
 /chloroplast  
 /strain="Goffinet 3162 (ALTA)"  
 /db\_xref="taxon:61563"  
 <1..>1320  
 /gene="rbcL"  
 <1..>1320  
 /gene="rbcL"  
 /codon\_start=1  
 /product="ribulose-bisphosphate carboxylase large subunit"  
 /protein\_id="AAC96050.1"  
 /db\_xref="PID:94009400"  
 /db\_xref="GI:4009400"  
 /translation="AGFRAGVKDYRLTYYPDPYQTKDTILIAFRMTPOGCPVPAEAG  
 AAVAESSTGTWTWTDGLTSLDRYRCYDEAVGEGDNOYIAYVAPDLPEBS  
 VTNLFTSGVNGVEFGKALRALRLDLRLPAYSKTFOGPPHIGVDRDLKRYRPL  
 GCTIKPKGLSANKYGRAVYECLAGGLDFTKDDNVNVSOPPRMRDPLFCALIVYS  
 OAFETGKIGLAKNTACGCEPMKRAPEAREICAPYIMHOVLTGCFPANTLAVYCD  
 NGLLHTRAHNAVITDKKNHGMHFRVLAKRLSGGDHITHAGTVAGKLEERVYTLG  
 FVDLLRDYIEKDRSRGTYFTQDVNSLPGLPVASGGIHWHPALTEIFEDSDVLP  
 GCGTILGHMGWNPAGVANNRVALLKCAVARNRGRLAREGNDVIREATK"

BASE COUNT 392 a 211 c 279 g 438 t  
 ORIGIN

Query Match 82.9%; Score 17.4; DB 8; Length 1320;  
 Best Local Similarity 94.7%; Pred. No. 1.2e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 aatttcggagatgatc 21  
 ||||||||||||||||  
 Db 1145 AAATTTTCGAGATGATTC 1163

RESULT 13  
 AF005539  
 LOCUS AF005539 1320 bp DNA PLN 17-DEC-1998  
 DEFINITION Ulota obtusiuscula ribulose-bisphosphate carboxylase large subunit  
 (rbcL) gene, chloroplast gene encoding chloroplast protein, partial  
 cds.  
 ACCESSION AF005539  
 NID 94009405  
 VERSION AF005539.1 GI:4009405  
 KEYWORDS  
 SOURCE  
 ORGANISM Ulota obtusiuscula.  
 Chloroplast Ulota obtusiuscula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryopsida;  
 Bryidae; Orthotrichales; Orthotrichaceae; Ulota.  
 REFERENCE  
 AUTHORS Goffinet,B., Bayer,R.J. and Vitt,D.H.  
 TITLE 1 (bases 1 to 1320)  
 JOURNAL Goffinet,B., Bayer,R.J. and Vitt,D.H.  
 REFERENCE Circumscription and phylogeny of the orthotrichales (Bryopsida)  
 Am. J. Bot. 85, 1324-1337 (1998)  
 2 (bases 1 to 1320)  
 AUTHORS Goffinet,B.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAY-1997) Department of Botany, Duke University,

P.O.Box 90339, Durham, NC 27708

FEATURES  
source  
Location/Qualifiers  
1. .1320  
/organism="Ulocha obtusiuscula"  
/chloroplast  
/strain="Goffinet 3161 (ALTA)"  
/db\_xref="taxon:61579"  
<1. .>1320  
/gene="rbcl"  
<1. .>1320  
/gene="rbcl"  
/codon\_start=1  
/product="ribulose-bisphosphate carboxylase large subunit"  
/protein\_id="AAC96053.1"  
/db\_xref="PID:94009406"  
/db\_xref="GI:4009406"  
/translation="VGFKAGVADRLTYTPDYOTKENDILAAFRMPQPGVPAEAG  
AAVAESSTGTWTTWDGLTSLDRYGRCDLAVPEDNQYIAVAYPLDLEEGS  
VTNLETSIVGNVFGFKALRALRLDRLIPPAVSKTFQGPPIGIVERRKLNKGRPL  
GCTIKPKIGLSAKNYGRAVYECLEGLDFTDDENVNSQPPMRDRFLFCALAIYS  
QAETGEIKGHYLNATAGTCEEMKRAECARELGMPIVHMDYLTGFTANTLAIYCRD  
NGLLHIRAHAAVAVIDROKNHGMHFRVLAARLRLSGDHIHAGTVGKLEGEROYTLG  
FVLDLDDYIEKDRSGIETFDQVNSLPGVLPVASSGITHVHXPALTEIFDGDVLOP  
GGTILGHMGNAPGAVANRVALEACVQARNRGNRLARNGNEVIREATK"

BASE COUNT  
394 a 209 c 280 g 436 t 1 others

ORIGIN

Query Match  
Best Local Similarity 94.7%; Score 17.4; DB 8; Length 1320;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 aattttcgagatgattc 21  
|||  
DB 1145 AATTTTCGAGATGATTTC 1163

RESULT 14  
AF005540 1320 bp DNA PLN 17-DEC-1998  
LOCUS Ulocha ribulose-bisphosphate carboxylase large subunit (rbcl)  
DEFINITION gene, chloroplast gene encoding chloroplast protein, partial cds.  
ACCESSION AF005540  
NID 94009407  
VERSION AF005540.1 GI:4009407  
KEYWORDS  
SOURCE  
ORGANISM  
Ulocha lutea.  
Chloroplast Ulocha lutea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryopsida;  
Bryidae; Orthotrichales; Orthotrichaceae; Ulocha.  
REFERENCE  
AUTHORS Goffinet B., Bayer R.J. and Vitt D.H.  
TITLE Circumscriton and phylogeny of the Orthotrichales (Bryopsida)  
JOURNAL Am. J. Bot. 85, 1324-1337 (1998)  
REFERENCE  
AUTHORS Goffinet B.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAY-1997) Department of Botany, Duke University,  
P.O.Box 90339, Durham, NC 27708  
FEATURES  
source  
Location/Qualifiers  
1. .1320  
/organism="Ulocha lutea"  
/chloroplast  
/strain="Pile 8042 (ALTA)"  
/db\_xref="taxon:61578"  
<1. .>1320  
/gene="rbcl"  
<1. .>1320  
/gene="rbcl"  
/codon\_start=1  
/product="ribulose-bisphosphate carboxylase large subunit"  
/protein\_id="AAC96054.1"

/db\_xref="PID:94009408"  
/db\_xref="GI:4009408"  
/translation="VGFKAGVADRLTYTPDYOTKENDILAAFRMPQPGVPAEAG  
AAVAESSTGTWTTWDGLTSLDRYGRCDLAVPEDNQYIAVAYPLDLEEGS  
VTNLETSIVGNVFGFKALRALRLDRLIPPAVSKTFQGPPIGIVERRKLNKGRPL  
GCTIKPKIGLSAKNYGRAVYECLEGLDFTDDENVNSQPPMRDRFLFCALAIYS  
QAETGEIKGHYLNATAGTCEEMKRAECARELGMPIVHMDYLTGFTANTLAIYCRD  
NGLLHIRAHAAVAVIDROKNHGMHFRVLAARLRLSGDHIHAGTVGKLEGEROYTLG  
FVLDLDDYIEKDRSGIETFDQVNSLPGVLPVASSGITHVHXPALTEIFDGDVLOP  
GGTILGHMGNAPGAVANRVALEACVQARNRGNRLARNGNEVIREATK"

BASE COUNT  
395 a 209 c 277 g 438 t 1 others

ORIGIN

Query Match  
Best Local Similarity 94.7%; Score 17.4; DB 8; Length 1320;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 aattttcgagatgattc 21  
|||  
DB 1145 AATTTTCGAGATGATTTC 1163

RESULT 15  
AF005541 1320 bp DNA PLN 17-DEC-1998  
LOCUS Bryodioxonia perichaetialis  
DEFINITION subunit (rbcl) gene, chloroplast gene encoding chloroplast protein,  
partial cds.  
ACCESSION AF005541  
NID 94009409  
VERSION AF005541.1 GI:4009409  
KEYWORDS  
SOURCE  
ORGANISM  
Bryodioxonia perichaetialis.  
Chloroplast Bryodioxonia perichaetialis  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryopsida;  
Bryidae; Orthotrichales; Orthotrichaceae; Bryodioxonia.  
REFERENCE  
AUTHORS Goffinet B., Bayer R.J. and Vitt D.H.  
TITLE Circumscriton and phylogeny of the Orthotrichales (Bryopsida)  
JOURNAL Am. J. Bot. 85, 1324-1337 (1998)  
REFERENCE  
AUTHORS Goffinet B.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAY-1997) Department of Botany, Duke University,  
P.O.Box 90339, Durham, NC 27708  
FEATURES  
source  
Location/Qualifiers  
1. .1320  
/organism="Bryodioxonia perichaetialis"  
/chloroplast  
/strain="Pile 8083 (ALTA)"  
/db\_xref="taxon:61547"  
<1. .>1320  
/gene="rbcl"  
<1. .>1320  
/gene="rbcl"  
/codon\_start=1  
/product="ribulose-bisphosphate carboxylase large subunit"  
/protein\_id="AAC96055.1"  
/db\_xref="PID:94009410"  
/db\_xref="GI:4009410"  
/translation="VGFKAGVADRLTYTPDYOTKENDILAAFRMPQPGVPAEAG  
AAVAESSTGTWTTWDGLTSLDRYGRCDLAVPEDNQYIAVAYPLDLEEGS  
VTNLETSIVGNVFGFKALRALRLDRLIPPAVSKTFQGPPIGIVERRKLNKGRPL  
GCTIKPKIGLSAKNYGRAVYECLEGLDFTDDENVNSQPPMRDRFLFCALAIYS  
QAETGEIKGHYLNATAGTCEEMKRAECARELGMPIVHMDYLTGFTANTLAIYCRD  
NGLLHIRAHAAVAVIDROKNHGMHFRVLAARLRLSGDHIHAGTVGKLEGEROYTLG  
FVLDLDDYIEKDRSGIETFDQVNSLPGVLPVASSGITHVHXPALTEIFDGDVLOP  
GGTILGHMGNAPGAVANRVALEACVQARNRGNRLARNGNEVIREATK"

BASE COUNT  
394 a 209 c 277 g 439 t 1 others

ORIGIN

Thu Aug 19 10:47:07 1999

us-09-004-395-6.rge

Page 10

	Query Match	Similarity	Score	DB	length
	Best Local	Similarity	94.7%	Pred. No. 1	2e20:
Matches	18;	Conservative	0;	Mismatches	1;
				Indels	0;
				Gaps	0;
Qy	3	aattttcgagatgattc	21		
	11				
	11				
Db	1145	AAATTTCGAGATGATTTC	1163		

Search completed: August 18, 1999, 17:17:48  
Job time: 3514 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 17:18:42 ; Search time 148.09 seconds  
(Without alignments)  
35.479 Million cell updates/sec

Title: US-09-004-395-6

Sequence: 1 ctaatttcgagatgattc 21

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database: N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	110000	1 X20248_01	Continuation (2 of
2	21	100.0	110000	1 X20248_02	Continuation (3 of
3	16.8	80.0	18613	1 V74423	Staphylococcus aur
4	16.4	78.1	900	1 X14196	H. pylori GHPO 903
5	16.2	77.1	1536	1 O47789	SNP gene. Yeast S
6	16.2	77.1	20633	1 X13213	Enterococcus faeca
7	15.8	75.2	12923	1 N90338	Sequence of human
8	15.8	75.2	23736	1 V57903	Hereditary haemoch
9	15.8	75.2	7696	1 X13474	Enterococcus faeca
10	15.8	75.2	32768	1 X13037	Enterococcus faeca
11	15.8	75.2	110000	1 X20248_02	Continuation (3 of
12	15.4	73.3	10320	1 T74666	Full-length utroph
13	15.4	73.3	346	1 V27287	Arabidopsis thalia
14	15.4	73.3	17764	1 X13238	Enterococcus faeca
15	15.2	72.4	1275	1 T68023	H. pylori secreted
16	15.2	72.4	2484	1 V52367	Streptococcus pneu
17	15.2	72.4	10659	1 V52190	Streptococcus pneu
18	15.2	72.4	19446	1 V52184	Streptococcus pneu
19	15.2	72.4	686	1 T98735	DNA encoding a S.
20	15.2	72.4	1555	1 X14096	H. pylori GHPO 132
21	15.2	72.4	1575	1 X05954	Heat-resistant gly
22	15.2	72.4	4104	1 X07356	Arabidopsis sterol
23	15.2	72.4	1242	1 X30451	H. pylori secreted
24	14.8	70.5	4293	1 O05366	Sequence of plasmid
25	14.8	70.5	1905	1 O05798	Synthetic B.t.k. p
26	14.8	70.5	4140	1 O30109	Encodes acid alpha
27	14.8	70.5	294	1 O60293	Human brain expres
28	14.8	70.5	2487	1 O45332	Human epidermal su
29	14.8	70.5	5261	1 O99602	Human Ah receptor
30	14.8	70.5	3300	1 T41617	Adzuki bean endo-x
31	14.8	70.5	19932	1 T46159	H. pylori secreted
32	14.8	70.5	1257	1 T68125	H. pylori secreted
33	14.8	70.5	1257	1 T67564	Human Ah-receptor
34	14.8	70.5	5261	1 T85436	DNA for epidermal
35	14.8	70.5	2486	1 T95887	Insecticidal gene
36	14.8	70.5	1919	1 V00396	H. pylori ORF 06gp
37	14.8	70.5	726	1 V24850	H. pylori secreted
38	14.8	70.5	714	1 V24949	Probe (104) for ml
39	14.8	70.5	307	1 V20203	Streptococcus pneu
40	14.8	70.5	6211	1 V52141	Staphylococcus aur
41	14.8	70.5	10953	1 V74373	Continuation (8 of
42	14.8	70.5	110000	1 X20248_07	Continuation (9 of
43	14.8	70.5	110000	1 X20248_08	

## ALIGNMENTS

44 14.6 69.5 134525 1 004525  
45 14.6 69.5 1200 1 V74974

Total base sequenc  
Staphylococcus aur

RESULT 1  
X20248\_01  
Continuation (2 of 10) of X20248 from base 100001 (Borrelia burgdorferi polynucleotide  
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
WP Fragment Name Begin End  
WP X20248\_00 1 110000  
WP X20248\_01 100001 210000  
WP X20248\_02 200001 310000  
WP X20248\_03 300001 410000  
WP X20248\_04 400001 510000  
WP X20248\_05 500001 610000  
WP X20248\_06 600001 710000  
WP X20248\_07 700001 810000  
WP X20248\_08 800001 910000  
WP X20248\_09 900001 910715

Query Match 100.0%; Score 21; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctaatttcgagatgattc 21

Db 104115 CTAATTTTCGAGATGATTC 104135

RESULT 2  
X20248\_02  
Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide  
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
WP Fragment Name Begin End  
WP X20248\_00 1 110000  
WP X20248\_01 100001 210000  
WP X20248\_02 200001 310000  
WP X20248\_03 300001 410000  
WP X20248\_04 400001 510000  
WP X20248\_05 500001 610000  
WP X20248\_06 600001 710000  
WP X20248\_07 700001 810000  
WP X20248\_08 800001 910000  
WP X20248\_09 900001 910715

Query Match 100.0%; Score 21; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctaatttcgagatgattc 21

Db 4115 CTAATTTTCGAGATGATTC 4135

RESULT 3  
V74423/c  
ID V74423 standard: DNA; 18613 BP.  
AC V74423;  
DT 16-MAR-1999 (first entry)  
DE Staphylococcus aureus contig SEQ ID #112.  
KW Computer readable medium; vaccine; S. aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
OS Staphylococcus aureus.  
FH Key Location/Qualifiers  
FT misc\_feature 901..960  
FT /tag= a

[illegible]

```

CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 18613 BP; 5990 A; 2753 C; 3477 G; 5782 T;

Query Match 80.0%; Score 16.8; DB 1; Length 18613;
Best Local Similarity 90.0%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taattttcggagatgattc 21
||| ||| ||| ||| ||| |||
DB 12855 TAAATTTTGGAGATGATTC 12836

RESULT 4
X14196/c X14196 standard; DNA; 900 BP.
AC ID X14196;
DT 31-MAR-1999 (first entry)
DE H. pylori GHP0 903 gene.
KW GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT CDS 59..781
FT /tag=a

PN WO9843478-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06571.
PR 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PI (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PA Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WP1; 98-542293/46.
DR P-PSDB; W98477.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 1; Page 859-860; 2054pp; English.
PS This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHP0 protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 900 BP; 308 A; 183 C; 167 G; 242 T;

Query Match 78.1%; Score 16.4; DB 1; Length 900;
Best Local Similarity 94.4%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 aattttcggagatgattc 20
||||| ||| ||| ||| |||
DB 570 AATTTTTCGGCGATGATTC 553

```



```
RESULT 5
ID Q47789 standard; DNA; 1536 BP.
AC Q47789;
DE 02-MAR-1994 (first entry)
KW SHP gene.
KW yeast; SNF1; homologous protein; SHP; Nicotiana tabacum; Invertase;
SS.
OS yeast sp.
PN J05199884-A.
PD 10-AUG-1993.
PF 13-DEC-1991; 330417.
PR 13-DEC-1991; JP-330417.
PA (SUMO) SUMITOMO CHEM CO LTD.
DR WPI: 93-284682/36.
P-PSDB: R40842.
PT yeast SNF1 homologous protein phosphorylase gene of Solanaceae
PT plants - useful for controlling expression of Invertase for the
PT increase of crop yield
PS Claim 1; Page 6; JP; Japanese.
CC This sequence represents the yeast SNF1 homologous protein gene
CC (SHP). The yeast SHP gene may be obtained from Nicotiana tabacum
CC and may be used to control expression of Invertase for the increase
CC of crop yield.
SQ Sequence 1536 BP; 437 A; 291 C; 382 G; 426 T;

Query Match 77.1%; Score 16.2; DB 1; Length 1536;
Best Local Similarity 85.7%; Pred. No. 31;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ctaatttcggagatgattc 21
    ||||| ||||| ||||| |||||
Db 1344 CCAATTCCTTGAGATGATTC 1364

RESULT 6
ID X13213/C
ID X13213 standard; DNA; 20633 BP.
AC X13213;
DE 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:276.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN M09850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI: 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1; Page 1284-1294; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
```

```
CC infection.
SQ Sequence 20633 BP; 5857 A; 4284 C; 3025 G; 7427 T;

Query Match 77.1%; Score 16.2; DB 1; Length 20633;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ctaatttcggagatgattc 21
    ||||| ||||| ||||| |||||
Db 10122 CTAATTTGCCGAGATGATTC 10102

RESULT 7
ID N90338/C
ID N90338 standard; cDNA; 12923 BP.
AC N90338;
DE 29-MAR-1992 (first entry)
DE Sequence of human muscular dystrophy (MD) cDNA.
KW dystrophin; muscular dystrophy; probe; diagnosis;
KW prenatal; heterozygote; gene therapy; genetic screening;
KW foetal screening; ss.
OS Homo sapiens.
FH key Location/Qualifiers
FT cds 209..12923
FT /*tag= a
PN M08906286-A.
PD 13-JUL-1989.
PF 16-DEC-1988; U04504.
PR 22-DEC-1987; US-136618.
PA (CHIL-) CHILDRENS MED CENT.
PI Kunkel LM, Monaco A, Hoffman EP, Koenig M;
DR WPI: 89-220587/30.
P-PSDB: P90373.
PT Muscular dystrophy gene - used for prepn. of probes, dystrophin
PT polypeptide and antibodies for diagnosis and therapy of muscular
PT dystrophy
PS Disclosure: Fig 5; 68pp; English.
CC The inventors claim an MD probe comprising a purified ss NA SQ which
CC hybridizes to at least a part of the MD gene; pure dystrophin (DS)
CC polypeptide, purified NA encoding DS and antibodies (Ab) to DS. The
CC probes are equal to or greater than 10b of one of 12 cDNA sequences
CC deposited as ATCC 58666-57677. The MD gene is human, or a murine Dmd
CC gene.
SQ Sequence 12923 BP; 4296 A; 2613 C; 2963 G; 3044 T;

Query Match 75.2%; Score 15.8; DB 1; Length 12923;
Best Local Similarity 89.5%; Pred. No. 58;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 aatttttcggagatgattc 21
    ||||| ||||| ||||| |||||
Db 3367 AATTTTCGAGATTATTC 3349

RESULT 8
ID V57903/C
ID V57903 standard; DNA; 237326 BP.
AC V57903;
DE 21-DEC-1998 (first entry)
DE Hereditary haemochromatosis subregion from an HH affected individual.
DE Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
KW diagnosis; iron metabolism; NPT3; NPT4; KOREC; BTF1; BTF2; BTF3;
KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW type 1 sodium transport gene; ss.
OS Homo sapiens.
PN M09814466-A1.
PD 09-APR-1998.
PF 30-SEP-1997; U17658.
PF 07-MAY-1997; US-852495.
PR 01-OCT-1996; US-724394.
PA (PROG-) PROGEANTOR INC.
```

PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,  
 PI Tsuchinashi Z, Wolff RK;  
 DR WPI: 98-240014/21.  
 CC Hereditary haemochromatosis gene products - used to develop products  
 PT for the diagnosis and treatment of hereditary disorders in iron  
 PT metabolism  
 PS Claim 1: Fig 9: 209pp; English.  
 CC The present invention describes hereditary haemochromatosis gene  
 CC products from the human haemochromatosis gene. The present sequence  
 CC represents a hereditary haemochromatosis subregion from an hereditary  
 CC haemochromatosis (HH) affected individual. Also described is a  
 CC method to determine the presence or absence of the common hereditary  
 CC haemochromatosis (HFE) gene mutation in an individual comprising:  
 CC (a) providing DNA or RNA from the individual; and (b) assessing the  
 CC DNA or RNA for the presence or absence of a haplotype or genotype where  
 CC the presence or absence of the haplotype genotype indicates the likely  
 CC presence of the HFE gene mutation in the genome of the individual. The  
 CC HFE gene sequences from the present invention can be used to develop  
 CC products for use in the diagnosis and treatment of HFE. The present  
 CC invention also describes BTF genes, which are homologues of the milk  
 CC protein butyrophilin (BT), and can be used in the production of agonists  
 CC and antagonists of BT function. Also described are: (1) a ROR1 gene  
 CC which can be used to develop products for the study, diagnosis and  
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes  
 CC which are homologues of a type 1 sodium transport gene, and can  
 CC similarly be used for hypophosphatemia.  
 SQ Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T;

Query Match 75.2%; Score 15.8; DB 1; Length 237326;  
 Best Local Similarity 89.5%; Pred. No. 72;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 taatttcgagatgatt 20  
 ||||| ||||| |||||  
 DB 210905 TATTTCGAGATGATT 210887

RESULT 9  
 ID X13474 standard; DNA: 7696 BP.  
 AC X13474:  
 DT 19-MAR-1999 (first entry)  
 DE Enterococcus faecalis genome contig SEQ ID NO:537.  
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;  
 KM vaccine; attenuation; computer readable medium; ds.  
 OS Enterococcus faecalis.  
 PN WO9850555-A2.  
 PD 12-NOV-1998.  
 PF 04-MAY-1998; 008985.  
 PR 14-NOV-1997; US-066009.  
 PR 06-MAY-1997; US-044031.  
 PR 16-MAY-1997; US-046655.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Dillon PU, Kunsch CA;  
 PI WPI: 99-045171/04.  
 PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
 PT - used to develop products for the detection of Enterococcus and for  
 PT use in vaccines for prevention or attenuation of Enterococcus  
 PT infection.  
 PS Claim 1: Page 1789-1793; 2084pp; English.  
 CC A computer readable medium has been developed which has recorded on it  
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
 CC X12938 to X13919 represent these nucleotide sequences which are primary  
 CC nucleotide sequences, also known as contigs. The computer-based system  
 CC can identify fragments of the Enterococcus faecalis genome with  
 CC commercial importance. The products can be used to detect the presence  
 CC of Enterococcus faecalis in samples. They can also be used for  
 CC diagnosing Enterococcal infection in an animal and monitoring  
 CC progression of disease, and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC can be used in vaccines to prevent or attenuate an Enterococcal  
 CC infection.  
 SQ Sequence 7696 BP; 2351 A; 1604 C; 1326 G; 2409 T;

Query Match 75.2%; Score 15.8; DB 1; Length 7696;  
 Best Local Similarity 89.5%; Pred. No. 56;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 aatttcgagatgattc 21  
 ||||| ||||| |||||  
 DB 7535 AATTTCGAGATGATTTC 7553

RESULT 10  
 ID X13037 standard; DNA: 32768 BP.  
 AC X13037:  
 DT 19-MAR-1999 (first entry)  
 DE Enterococcus faecalis genome contig SEQ ID NO:100.  
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;  
 KM vaccine; attenuation; computer readable medium; ds.  
 OS Enterococcus faecalis.  
 PN WO9850555-A2.  
 PD 12-NOV-1998.  
 PF 04-MAY-1998; 008985.  
 PR 14-NOV-1997; US-066009.  
 PR 06-MAY-1997; US-044031.  
 PR 16-MAY-1997; US-046655.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barash SC, Dillon PU, Kunsch CA;  
 PI WPI: 99-045171/04.  
 PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
 PT - used to develop products for the detection of Enterococcus and for  
 PT use in vaccines for prevention or attenuation of Enterococcus  
 PT infection.  
 PS Claim 1: Page 645-661; 2084pp; English.  
 CC A computer readable medium has been developed which has recorded on it  
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
 CC X12938 to X13919 represent these nucleotide sequences which are primary  
 CC nucleotide sequences, also known as contigs. The computer-based system  
 CC can identify fragments of the Enterococcus faecalis genome with  
 CC commercial importance. The products can be used to detect the presence  
 CC of Enterococcus faecalis in samples. They can also be used for  
 CC diagnosing Enterococcal infection in an animal and monitoring  
 CC progression of disease, and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
 CC another related organism, in vivo or in vitro. In particular the  
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an Enterococcal  
 CC infection.  
 SQ Sequence 32768 BP; 9063 A; 7056 C; 5292 G; 11350 T;

Query Match 75.2%; Score 15.8; DB 1; Length 32768;  
 Best Local Similarity 89.5%; Pred. No. 63;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 aatttcgagatgattc 21  
 ||||| ||||| |||||  
 DB 14783 AATTTCGAGATGATTTC 14765

RESULT 11  
 ID X20248\_02/c  
 DT Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide  
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248  
 WP Fragment Name Begin End  
 WP X20248\_00 1 110000  
 WP X20248\_01 100001 210000  
 WP X20248\_02 200001 310000  
 WP X20248\_03 300001 410000  
 WP X20248\_04 400001 510000



CC can identify fragments of the Enterococcus faecalis genome with  
 CC commercial importance. The products can be used to detect the presence  
 CC of Enterococcus faecalis in samples. They can also be used for  
 CC diagnosing Enterococcal infection in an animal and monitoring  
 CC progression of disease, and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
 CC another related organism, in vivo or in vitro. In particular the  
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an Enterococcal  
 CC infection.

SO Sequence 17764 BP; 5508 A; 3640 C; 2874 G; 5718 T;

Query Match 73.3%; Score 15.4; DB 1; Length 17764;  
 Best Local Similarity 94.1%; Pred. No. 93;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 attttcgagatgatt 20  
 ||||| ||||| |||||  
 Db 15742 ATTTTAGGAGATGAT 15726

## RESULT 15

ID T68023 standard; DNA: 1275 BP.  
 AC T68023;

DT 15-JUL-1997 (first entry)

DE H. pylori secreted or periplasmic protein ORF 079e11504orf2.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 identification; binding compound; bacterium; life cycle; activator;  
 bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;

KW diagnosis; ds.

OS Helicobacter pylori.

FH Key Location/Qualifiers

FT 1. 1275

FT cds /tag= a

FT /note= "no stop codon given"

PN W09640893-A1.

PD 19-DEC-1996.

PF 06-JUN-1996; U09122.

PR 07-JUN-1995; US-487032.

PR 01-APR-1996; US-630405.

PA (ASTR) ASTRA AB.

PI Berglindh OT, Smith D, Meligaard BL;

DR WPI: 97-052306/05.

P-PSDB: W20770.

PT Helicobacter pylori nucleic acid sequences and related

PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

PT infection, and to detect Helicobacter

PS Claim 23; Page 851; 1481pp; English.

CC This sequence encodes a H. pylori secreted or periplasmic protein.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors.

CC The genomic sequence of H. pylori (ATCC 55679) was determined from

CC overlapping contigs generated by mechanically shearing the bacterial

CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,

CC and the predicted coding regions defined by computer evaluation. To

CC identify likely H. pylori antigens for vaccine development, the amino

CC acid sequences predicted from various ORF were analysed for significant

CC homology to other known or exported membrane proteins. Having identified

CC and determined the sequences of interest, particular regions can be

CC isolated from H. pylori by PCR amplification for recombinant polypeptide

CC production, e.g. in E. coli hosts.

SO Sequence 1275 BP; 478 A; 226 C; 280 G; 291 T;

Query Match 72.4%; Score 15.2; DB 1; Length 1275;

Best Local Similarity 85.0%; Pred. No. 93;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ctaatttcgagatgatt 20  
 ||||| ||||| |||||

Db 1195 CTAATTTTGGGCGATGAT 1176

Search completed: August 18, 1999, 17:18:51  
 Job time: 3283 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 1999, 17:04:50 ; Search time 767.16 Seconds  
(without alignments)  
53.996 Million cell updates/sec

Title: us-09-004-395-6

Perfect score: 21  
Sequence: 1 ctaatttcgagatgattc 21

Scoring table: IDENTITY\_NUC

Searched: 2546578 segs, 986266752 residues

Database :

EST:  
1: em\_est1:  
2: em\_est2:  
3: em\_est3:  
4: em\_est4:  
5: em\_est5:  
6: em\_est6:  
7: em\_est7:  
8: em\_est8:  
9: em\_est9:  
10: em\_est10:  
11: em\_est11:  
12: em\_est12:  
13: em\_est13:  
14: em\_est14:  
15: em\_est15:  
16: em\_est16:  
17: em\_est17:  
18: em\_est18:  
19: em\_est19:  
20: em\_est20:  
21: em\_est21:  
22: em\_est22:  
23: em\_est23:  
24: em\_est24:  
25: em\_est25:  
26: em\_est26:  
27: em\_est27:  
28: em\_est28:  
29: em\_est29:  
30: em\_est30:  
31: em\_est31:  
32: em\_est32:  
33: em\_est33:  
34: em\_est34:  
35: em\_est35:  
36: em\_est36:  
37: em\_est37:  
38: em\_est38:  
39: em\_est39:  
40: em\_est40:  
41: em\_est41:  
42: em\_est42:  
43: em\_est43:  
44: em\_est44:  
45: em\_est45:  
46: em\_est46:  
47: em\_est47:  
48: em\_est48:  
49: em\_est49:  
50: em\_est50:  
51: em\_est51:  
52: em\_est52:  
53: em\_est53:

54: em\_est22:  
55: em\_est23:  
56: em\_est24:  
57: em\_est25:  
58: em\_est26:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17.8	84.8	585	39	C87660	C87660 C87660 Mous
2	16.8	80.0	300	35	C36783	C36783 C36783 Yuj1
3	16.2	77.1	441	29	AA184418	AA184418 m34h06.r
4	16.2	77.1	424	30	AA229217	AA229217 nc44b01.r
5	16.2	77.1	423	30	AA229337	AA229337 nc44b11.r
6	16.2	77.1	450	38	AA777778	AA777778 z195h12.s
7	16.2	77.1	426	41	AI033116	AI033116 o998h04.s
8	16.2	77.1	334	48	AI565372	AI565372 t173f05.x
9	16.2	77.1	493	51	AI723359	AI723359 ME000186.
10	16.2	76.2	398	24	H66575	H66575 yu16h03.r1
11	15.8	75.2	543	28	AA087198	AA087198 mk22f03.r
12	15.8	75.2	306	39	C24271	C24271 C24271 M1y8
13	15.8	75.2	312	39	C24276	C24276 C24276 M1y8
14	15.8	75.2	305	41	AI041524	AI041524 uc76e09.x
15	15.8	75.2	360	41	AI042694	AI042694 uc76e09.x
16	15.8	75.2	769	47	AI525474	AI525474 P1.1_3-E
17	15.8	75.2	364	48	AI579389	AI579389 UI-R-YO-V
18	15.8	75.2	431	49	AI621835	AI621835 486027B04
19	15.4	73.3	345	21	FI4032	FI4032 ARTS4954_Ve
20	15.4	73.3	428	22	R51895	R51895 y978d06.s1
21	15.4	73.3	441	23	D59634	D59634 HUM070A06B
22	15.4	73.3	453	25	W05124	W05124 za83h08.r1
23	15.4	73.3	449	26	W21025	W21025 zb55h12.r1
24	15.4	73.3	449	31	AA282333	AA282333 z112e11.r
25	15.4	73.3	340	32	AA354148	AA354148 EST62365
26	15.4	73.3	423	33	AA400732	AA400732 z709e10.r
27	15.4	73.3	463	38	AA808926	AA808926 nu29c01.r
28	15.4	73.3	654	40	C90900	C90900 C90900 Dict
29	15.4	73.3	714	42	AI124413	AI124413 SMOVL3CAN
30	15.4	73.3	471	42	AI146344	AI146344 qp41e07.x
31	15.4	73.3	474	42	AI148129	AI148129 qp43f12.x
32	15.4	73.3	359	46	AI422550	AI422550 CF86B03.X
33	15.2	72.4	245	20	D21613	D21613 MUS74B10 mo
34	15.2	72.4	138	20	D29292	D29292 HUMNK375 Hu
35	15.2	72.4	440	21	T86427	T86427 yd87d01.s1
36	15.2	72.4	305	21	T91811	T91811 ye02a01.s1
37	15.2	72.4	501	22	R55899	R55899 y992d01.r1
38	15.2	72.4	513	22	R56172	R56172 y992d01.r1
39	15.2	72.4	213	23	D60137	D60137 HUM087H10B
40	15.2	72.4	230	23	D61333	D61333 HUM180C04B
41	15.2	72.4	331	23	D61376	D61376 HUM184H11B
42	15.2	72.4	357	23	R85219	R85219 y041g10.s1
43	15.2	72.4	321	24	D61428	D61428 HUM400E07B
44	15.2	72.4	393	24	H84144	H84144 yf02g09.r1
45	15.2	72.4	597	49	AI630770	AI630770 tx53a10.x

# ALIGNMENTS

RESULT 1  
C87660  
LOCUS C87660 585 bp mRNA  
DEFINITION C87660 Mouse fertilized one-cell-embryo cDNA Mus musculus CDNA  
ACCESSION C87660  
NID g2919617  
VERSION C87660.1 GI:2919617

[illegible]

```

/clone="Yk467e5"
/clone_lib="Yuji Kohara unpublished cDNA"
/note="dev_stage=varied, sex=hermaphrodite male,
tissue_type=whole animal"
BASE COUNT      99 a      52 c      58 g      91 t
ORIGIN

Query Match      80.0%  Score 16.8;  DB 35;  Length 300;
Best Local Similarity 90.0%;  Pred. No. 1.4e+02;
Matches 18;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

QY      2  taatttttcgagatgattc 21
        ||||| ||||| ||||
DB      45  TAATTTTTCGAGATTAATTC 26

RESULT      3
Aa184418/c  Aa184418      441 bp      mRNA      EST      17-FEB-1997
LOCUS      mt34h06.1 Soares mouse 3nbms Mus musculus cDNA clone IMAGE:623003
DEFINITION  5', mRNA sequence.
Aa184418
G1768067
Aa184418.1  GI:1768067
EST.
house mouse.
SOURCE      Mus musculus
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE    Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS      1 (bases 1 to 441)
            Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE        The Mashu-HHMI Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      On Apr 14, 1993 this sequence version replaced g1:692827.

Contact: Marra M/Mouse EST Project
Mashu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel.: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LANT ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:383827
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 439.
Location/Qualifiers
1..441
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="623003"
/clone_lib="Soares mouse 3nbms"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer [5',
TGTTCACCAATCTGAAATGAGGAGCGCGCGCGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT733 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Palma Bonaldo."

```



0y	1	ctaatcttcggagatgattc	21	
			1	
Db	132	CTAATTTTACGACAGATTC	152	
RESULT	6			
LOCUS	AA777778/c			
DEFINITION	AA777777	450 bp	MRNA	EST
	2195H12.S1 Soares fetal liver spleen_1M15.S1			05-FEB-1998
	clone IMAGE:440583 3'			Homo sapiens cDNA
ACCESSION	AA777778			
NID	92837257			
VERSION	AA777778.1	GI:2837257		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
	Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
REFERENCE	1 (bases 1 to 450)			
AUTHORS	Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,			
	Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,			
	Martin,J., Moore,B., Schellenberg,K., Stepien,M., Tan,F.,			
	Reising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.			
TITLE	Washu-NCT human EST Project			
JOURNAL	Unpublished (1997)			
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2152720.			

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [estewatson.wustl.edu](mailto:estewatson.wustl.edu)  
This clone is available royalty-free through LINT; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: 40m3 fwd. Et from Amersham  
High quality sequence stop: 458.

FEATURES	Location/Qualifiers
SOURCE	1. .450

```

/organism="Homo sapiens"
/db_xref="GDB:1352840"
/db_xref="taxon:9606"
/map="1/q21"
/clone="IMAGE:448583"
/clone_lib="Soares_fetal_liver_spleen_1NFIS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI
This is a subtracted version of the original Soares fetal
liver spleen 1NFIS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
AAGCGAGAGATTATATTAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

```

	Query Match	77.1%	Score 16.2	DB 38	Length 450
Best Local Similarity	85.7%	Pred. No. 2	7e+02		
Matches 18	Conservative 0	Mismatches 3	Indels 0	Gaps 0	
QY	1 ctaatttcgagatgattc	21			
Db	245 CTCATTTCCTGAGATGATTC	225			

## RESULT 7

AI033116/c	AI033116	426 bp	mRNA	EST	28-AUG-1998
LOCUS	o98bh04.s1	Soares fetal_liver_spleen_1MFLS_S1	Homo sapiens	CDNA	
DEFINITION	clone IMAGE:1654903 3'	mRNA sequence.			
ACCESSION	AI033116				
NID	93254069				
VERSION	AI033116.1	GI:3254069			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1 (bases 1 to 426)				
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
JOURNAL	Tumor Gene Index				
COMMENT	Unpublished (1997)				
	On Jan 19, 1998 this sequence version replaced gi:2282244.				

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
This clone is available royalty-free through LINT ; contact the  
IMAGE Consortium ([info@image.lni.gov](mailto:info@image.lni.gov)) for further information.  
Insert length: 735 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 426.

## FEATURES

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1654903"
/clone_lib="Scorers_fetal_liver_spleen_INFLS_S1"
/sex="male"
/sex="male"
/der_stage="20 week post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pTR73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subcloned version of the Original Scores paired
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer 15',
AATCGAAGAATTTTATTAAGATCTCTTTTTTTTTTTTTTTT 3',
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pTR73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fátima Bonafide."

```

BASE COUNT	109 a	101 c	78 g	138 t
ORIGIN				
Query Match	77.1%: Score 16.2; DB 41;			
Best Local Similarity	85.7%: Pred. No. 2.7e+02;			
Matches 18; Conservative	0; Mismatches 3; Indels 0; Gaps 0;			
QY	1 ctaattttcggagatgattc 21			
Db	11111111111111111111			
	288 CTCATTTCCTGAGATGATTG 268			

RESULT	8				
AI565372					
LOCUS	AI565372	334 bp	mRNA	EST	26-MAR-1999
DEFINITION	t13f05.x1 NCI_CGAP_Kid1				
	mRNA sequence.				
ACCESSION	AI565372				
	94523829				
VERSION	AI565372.1	GI:4523829			
KEYWORDS	EST.				
SOURCE	human.				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.



REFERENCE 1 (bases 1 to 334)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On May 18, 1998 this sequence version replaced gi:3136815.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.lnlnl.gov/db/rlp/image/image.html](http://www-bio.lnlnl.gov/db/rlp/image/image.html)

Seq primer: -40UP from Gldco  
 High quality sequence stop: 307.

#### FEATURES

source  
 1..334  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2137665"  
 /clone\_1lb="NCI CGAP\_Kid11"  
 /lab\_host="DH10B"  
 /note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 Plasmid DNA from the normalized library NCI-CGAP\_Kid3 was  
 prepared, and ss circles were made in vitro. Following RNP  
 purification, this DNA was used as tracer in a subtracive  
 hybridization reaction. The driver was PCR-amplified cDNAs  
 from a pool of 5,000 clones made from the same library  
 (cloneids 132376-132391, 145607-145675, and  
 150052-150285). Subtraction by Bento Soares and M.  
 Fatima Bonaldo."

BASE COUNT 103 a 66 c 52 g 113 t  
 ORIGIN

Query Match 77.1%; Score 16.2; DB 48; Length 334;  
 Best Local Similarity 85.7%; Pred. No. 2.7e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 9  
 LOCUS AT723359 493 bp mRNA EST 10-JUN-1999  
 DEFINITION ME000186. FOR Egg stage cDNA expression library in Lambda ZAPIT  
 Schistosoma mansoni cDNA 5' similar to P47826 60S ACIDIC RIBOSOMAL  
 PROTEIN P0, mRNA sequence.  
 AT723359  
 ACCESSION AT723359  
 MID g5041687  
 VERSION AT723359.1 GI:5041687  
 KEYWORDS EST.  
 SOURCE Schistosoma mansoni.  
 ORGANISM Schistosoma mansoni.  
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
 Strigoidae; Schistosomatidae; Schistosomatidae; Schistosoma.

REFERENCE 1 (bases 1 to 493)  
 AUTHORS Oliveira, G.C. and Baba, J.  
 TITLE Cataloguing Schistosoma mansoni genes with expressed sequence tags  
 JOURNAL Unpublished (1998)  
 COMMENT On Dec 20, 1995 this sequence version replaced gi:1133305.

Contact: Oliveira, Guilherme  
 Lab. Parasitologia Cel. e Mol.

Centro de Pesquisas Rene Rachou - FIOCRUZ  
 Av. Augusto de Lima 1715 Barro Preto, Belo Horizonte, MG, CEP  
 30190, Brazil  
 Tel: 55 31 2953566  
 Fax: 55 31 2952115  
 Email: oliveira@eneira.cpqrr.fiocruz.br  
 Insert Length: 493 Std Error: 0.00  
 Seq primer: M13 Reverse Universal Sequencing primer.  
 Location/Qualifiers  
 1..493  
 /organism="Schistosoma mansoni"  
 /strain="Le"  
 /db\_xref="taxon:6183"  
 /map="17q21"  
 /clone\_1lb="Egg stage cDNA expression library in Lambda  
 ZAPIT"  
 /sex="Mixed"  
 /dev\_stage="Egg"  
 /note="Vector: Uni-Zap XR vector, Stratagene (pBluescript  
 SK+); Site\_1: EcoRI; Site\_2: XhoI; mRNA was extracted from  
 eggs and the library was constructed and excised  
 according to the manufacturer's instructions."

BASE COUNT 140 a 109 c 111 g 133 t  
 ORIGIN

Query Match 77.1%; Score 16.2; DB 51; Length 493;  
 Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 10  
 LOCUS H66575 398 bp mRNA EST 18-OCT-1995  
 DEFINITION yul6n03.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone  
 IMAGE:234005 5', mRNA sequence.  
 H66575  
 ACCESSION H66575  
 MID g1025315  
 VERSION H66575.1 GI:1025315  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 398)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Kohling, T., Soares, M., Tan, F.,  
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.  
 TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT On May 9, 1995 this sequence version replaced gi:802262.

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Insert Size: 1105  
 High quality sequence stops: 332  
 Source: IMAGE Consortium, LNL  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([infoimage.lnlnl.gov](http://infoimage.lnlnl.gov)) for further information.  
 Insert Length: 1105 Std Error: 0.00  
 Seq primer: M13RP1  
 High quality sequence stop: 332.  
 Location/Qualifiers

#### FEATURES

**Source**

```

source
1. .398
/organism="Homo sapiens"
/db_xref="GDB:3787182"
/db_xref="taxon:9606"
/clone="IMAGE:234005"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
15' AACGGAAGAATATATTAAGAATCTTTTTTTTTTTTTTTT 3',
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
94 a 85 c 86 g 132 t 1 others

```

Query Match	76.2%;	Score 16;	DB 24;	Length 398;
Best Local Similarity	100.0%;	Pred. No. 3.3e+02;		
Matches 16;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	5	ttttcggagatgatt	20
Db	132	TTTTTCGGAGATGATT	117

Db 132 TTTTTCGAGATGATT 117

RESULT	11
AA087198/c	
LOCUS	543 bp mRNA
DEFINITION	m222603.r1 Soares mouse p3MNF19.5 Mus musculus cDNA clone
ACCESSION	AA087198 IMAGE:493661 5', mRNA sequence.
NID	AA087198
VERSION	G1630409
KEYWORDS	AA087198.1 GI:1630409
SOURCE	EST.
ORGANISM	house mouse. Mus musculus

REFERENCE  
1 (bases 1 to 543)  
AUTHORS  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

TITLE The WashU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1294047

Contact: Maria M/Mouse EST Project  
 WashU-RHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LINT; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:297109  
 Putative full length read  
 vector to vector length is 609  
 Seq primer: 28M13 rev2 from Amersham  
 High quality sequence stop: 466.

```
FEATURES
source      Location/Qualifiers
1. 543
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="IMAGE:493661"
```

```

/clone_11b-"Soares mouse p3Mm19.5"
/dey_stgae="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGGAGCGCGCGGATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
```

Query Match	75.28;	Score 15.8;	DB 28;	Length 543;
Best Local Similarity	89.58;	Prod. No. 4.1e+02;		
Matches 17; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	3	aattttcggagatgattc	21
Db	543	acttttcagacatgatc	525

RESULT 12  
C24271/c

C24271/c

ACCESSION	unshlu cDNA clone pCMFR1804.93, mRNA sequence.
NID	C24271
VERSION	G2104068
KEYWORDS	G24271.1 GI:2104068
SOURCE	Esp.
ORGANISM	Citrus unshlu.
	Citrus unshlu
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euarhydropytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Sapindales; Rutaceae; Citrus. 1 (bases 1 to 306)
REFERENCE	

development phase  
J. Amer. Soc. Hort. Sci. 122, 808-812 (1997)  
On Jan 14, 1998 this sequence version replaced gl:1797773

Contact: Mitsuo Omura  
 Department of Citriculture  
 National Institute of Fruit Tree Science, Okitsu  
 Okitsu, Shizuoka, Shizuoka 424-02, Japan  
 Tel.: +81-543-69-2111  
 Fax: +81-543-69-2115  
 Email: om9330@okt.affrc.go.jp  
 PROJECT = Citrus genome Analysis.

FEATURES	source
Location/Qualifiers	
1. .306	
/organism="Citrus unshiu"	
/strain="Miyagawa-wase satsuma mandarin"	
/db_xref="taxon:55188"	
/map="15"	
/clone="pCMFR1804.93"	
/clone_1lb="Miyagawa-wase satsuma mandarin orange (M.Omura)"	
/tissue_type="juice sac and pulp segment"	
/dev_stage="rapid developing stage"	
70 a	93 c 84 g 59 t

ORIGIN	
Query Match	75.2%; Score 15.8; DB 39; Length 306;
Best Local Similarity	89.5%; Pred. No. 4.2e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 aatttcggagatgac 21  
 || ||||| ||||| |||||  
 Db 73 AAATTTGCGAGATCATTC 55

RESULT 13  
 LOCUS C24276/c 312 bp mRNA EST 29-JUN-1998  
 DEFINITION C24276 Miyagawa-wase satsuma mandarin orange (M.Omura) Citrus  
 unshu cDNA clone pCMFR1804.116, mRNA sequence.

ACCESSION C24276  
 MID 92104073  
 VERSION C24276.1 GI:2104073  
 KEYWORDS EST.  
 SOURCE Citrus unshu.  
 ORGANISM Citrus unshu.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicots; Rosidae; eurosids II; Sapindales; Rutaceae; Citrus.

REFERENCE  
 AUTHORS Hisada,S., Akihama,T., Endo,T., Moriguchi,T. and Omura,M.  
 TITLE Expressed sequence tags of Citrus fruit during rapid cell  
 development phase  
 JOURNAL J. Amer. Soc. Hort. Sci. 122, 808-812 (1997)  
 COMMENT On Jan 14, 1998 this sequence version replaced g1:187535.

FEATURES  
 source  
 1. 312  
 /organism="Citrus unshu"  
 /strain="Miyagawa-wase satsuma mandarin"  
 /db\_xref="taxon:55188"  
 /clone="pCMFR1804.116"  
 /clone\_id="Miyagawa-wase satsuma mandarin orange  
 (M.Omura)"  
 /tissue\_type="juice sac and pulp segment"  
 /dev\_stage="rapid developing stage"

BASE COUNT 73 a 94 c 86 g 59 t

ORIGIN

Query Match 75.2%; Score 15.8; DB 39; Length 312;  
 Best Local Similarity 89.5%; Pred. No. 4.2e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 aatttcggagatgac 21  
 || ||||| ||||| |||||  
 Db 73 AAATTTGCGAGATCATTC 55

RESULT 14  
 LOCUS A1041524 305 bp mRNA EST 30-JUN-1998  
 DEFINITION 3', mRNA sequence.  
 A1041524  
 A1041524  
 93280718  
 VERSION A1041524.1 GI:3280718  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
 1 (bases 1 to 305)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 19, 1998 this sequence version replaced g1:2150285.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bdip/image/image.html](http://www.bio.llnl.gov/bdip/image/image.html)

Seq primer: -40m13 fwd, ET from Amersham.  
 Location/Qualifiers  
 1. 305  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1643800"  
 /clone\_id="Soares\_testis\_NHT"  
 /sex="male"  
 /lab\_host="DH10B"  
 /note="Vector: pTR73D-Pac (Pharmacia) with a modified  
 polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech  
 Laboratories, Inc., and primed with a Not I - oligo(dT)  
 primer [5'.  
 TGTACCAATCTGAGTGGAGCGGCCGCCCAATTTTCTTTTCTTTT 3'].  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pTR73 vector. Library  
 went through one round of normalization to CoTs, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 86 a 62 c 68 g 89 t

ORIGIN

Query Match 75.2%; Score 15.8; DB 41; Length 305;  
 Best Local Similarity 89.5%; Pred. No. 4.2e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctatttcggagatgat 19  
 ||||| ||||| ||||| |||||  
 Db 30 CTAAGTTTCTGAGATGAT 48

RESULT 15  
 LOCUS A1042694 360 bp mRNA EST 01-JUL-1998  
 DEFINITION uc76e09.x1 Sugano mouse liver m1a Mus musculus cDNA clone  
 IMAGE:1431592 3', mRNA sequence.  
 A1042694  
 A1042694  
 93285954  
 VERSION A1042694.1 GI:3285954  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1 (bases 1 to 360)  
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

TITLE The WashU-HMNI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT On Jan 19, 1998 this sequence version replaced g1:2153086.

Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:915660

Seq primer: custom primer used.

## FEATURES

## Source

Location/Qualifiers  
 1..360  
 /organism="Mus musculus"  
 /strain="C57BL"  
 /db\_xref="taxon:10090"  
 /map="17q21"  
 /clone="IMAGE:1431592"  
 /clone\_lib="Sugano mouse liver ml1a"  
 /sex="female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: liver; Vector: pME18s-FL3; Site\_1: DraIII  
 (CACGCTGTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA  
 was primed with an oligo(dT) primer  
 [ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was  
 ligated to a DraIII adaptor (TGTGGCCCTACTG), digested  
 and cloned into distinct DraIII sites of the pME18s-FL3  
 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should  
 be used to isolate the cDNA insert. Size selection was  
 performed to exclude fragments <1.5kb. Library  
 constructed by Dr. Sumio Sugano (University of Tokyo  
 Institute of Medical Science). Custom primers for  
 sequencing: 5' end primer CTTCGCTCTTAAGCTGCG and 3' end  
 primer CGACCTGCAGCTCGACACA."  
 BASE COUNT 91 a 76 c 75 g 118 t  
 ORIGIN

## Query Match

Best Local Similarity 75.2%; Score 15.8; DB 41; Length 360;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctaatttcgagatgat 19

Db 44 CTAAGTTTCTGAGATGAT 62

Search completed: August 18, 1999, 17:04:54  
 Job time: 2800 sec

RESULT 2  
ID 051612 PRELIMINARY: PRT: 344 AA.  
AC 051612; (Tremblrel. 06, Created)  
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
DT 01-JUN-1998 (Tremblrel. 08, Last annotation update)  
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
DE FLAGELLAR FILAMENT OUTER LAYER PROTEIN (FLA).  
GN BB0668.  
OS Bacteriella burgdorferi (Lyme disease spirochete).  
OC Bacteriella; Spirochaetales; Spirochaetaceae; Borrelia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 35210 / B31;  
RX MEDLINE: 98065943.  
RA FRASER C.M., CASSENS S., HUANG M.M., SUTTON G.G., CLAYTON R.A.,  
RA LATHIGR R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,  
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,  
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,  
RA VAN VUOT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,  
RA UTTERBACK T., MATTHEY L., McDONALD L., ARTICH P., BOWMAN C.,  
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,  
RA SMITH H.O., VENTER J.C.;  
RT "Genomic sequence of a Lyme disease spirochete, Borrelia  
burgdorferi."  
RL Nature 390:580-586(1997).  
DR EMBL: A800168; AAC67025.1; .  
DR TIGR: BB0668; .  
KM Flagella.  
SQ SEQUENCE 344 AA; 38834 MW; 6CDD00CC1 CRC32;

Query Match 99.0%; Score 1701; DB 2; Length 344;  
Best Local Similarity 98.8%; Pred. No. 1.4e-109;  
Matches 337; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKRKASLFLFLSTLVFAOETDGLAEGSKRAPEGLVLDPAELARDPSSRLDLTNYVD 60  
DB 4 MKRKASLFLFLSTLVFAOETDGLAEGSKRAPEGLVLDPAELARDPSSRLDLTNYVD 63  
QY 61 YVTSAGSIVKPEDVNDLGINNSVLLTPSARLOAYKNSVAPAYKSESKRYAGDTI 120  
DB 64 YVTSAGSIVKPEDVNDLGINNSVLLTPSARLOAYKNSVAPAYKSESKRYAGDTI 123  
QY 121 LQVRVLFPSYSSSAMIMPFFKIPFYSGSGNQFLGKLIDNKTKEIKTSVYSLGEYI 180  
DB 124 LQVRVLFPSYSSSAMIMPFFKIPFYSGSGNQFLGKLIDNKTKEIKTSVYSLGEYI 183  
QY 181 DLEVFEDNMGMAYSMGLTKFKGNADLIWNSPNTIPNISSRIITKDDVPNTPLASSKR 240  
DB 184 DLEVFEDNMGMAYSMGLTKFKGNADLIWNSPNTIPNISSRIITKDDVPNTPLASSKR 243  
QY 241 FNAFRVSKSHSKVKNFIFVVKDLRYLKYDLSYSDSDISESVFYETSGTESLRK 300  
DB 244 FNAFRVSKSHSKVKNFIFVVKDLRYLKYDLSYSDSDISESVFYETSGTESLRK 303  
QY 301 AHETFRVLTREKISIAEGSFONFEKIESERPESSPRN 341  
DB 304 AHETFRVLTREKISIAEGSFONFEKIESERPESSPRN 344

RESULT 3  
ID 044876 PRELIMINARY: PRT: 66 AA.  
AC 044876;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)  
DE CHA GENE (FRAGMENT).  
OS Bacteriella burgdorferi (Lyme disease spirochete).  
OC Bacteriella; Spirochaetales; Spirochaetaceae; Borrelia.  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-212;  
RX MEDLINE: 98438936.  
RA OLD I.G., TRUEBA G.A., SAINT GIROIS I., JOHNSON R.C.;  
RT "A chea cheu operon in Borrelia burgdorferi, the agent of Lyme  
disease."  
RL Res. Microbiol. 148:191-200(1997).  
DR EMBL: X91907; CAA63001.1; .  
FT NON-TER 1 1  
SQ SEQUENCE 66 AA; 7577 MW; D66137D6 CRC32;

Query Match 19.0%; Score 326; DB 2; Length 66;  
Best Local Similarity 100.0%; Pred. No. 4e-16; Indels 0; Gaps 0;  
Matches 66; Conservative 0; Mismatches 0;

QY 276 DSDIDSESVFYETSGTESLRKKAHETFRVLTREKISIAEGSFONFEKIESERPE 335  
DB 1 DSDIDSESVFYETSGTESLRKKAHETFRVLTREKISIAEGSFONFEKIESERPE 60  
QY 336 ESSPRN 341  
DB 61 ESSPRN 66

RESULT 4  
ID 006691 PRELIMINARY: PRT: 243 AA.  
AC 006691;  
DT 01-JUL-1997 (Tremblrel. 04, Created)  
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)  
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)  
DE FLAA HOMOLOG-1.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA PORCELLA S.F., RADOLF J.D., NORGARD M.V.;  
RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U97363; AAB6367.1; .  
SQ SEQUENCE 243 AA; 27299 MW; 7CB7C276 CRC32;

Query Match 7.6%; Score 131.5; DB 2; Length 243;  
Best Local Similarity 22.5%; Pred. No. 0.057;  
Matches 56; Conservative 33; Mismatches 109; Indels 51; Gaps 8;

QY 48 PSSTRDLTNTVDY---YSGASGIVKPEDVNDLGINN---SVLLTPSARLOAYK 100  
DB 21 PLAAQRSAKVNVOAFIDFDGAS-----EDQGL-AMRAAGSKRTTKGFPILKTFEG 70  
QY 101 SYVAPAVYKSESKRYAGDTILGVRLFPYSOSSAMIMP-----PFKIPFYSGSGNQFL 155  
DB 71 MPQAVRMAGSMQCKREARFVGCEKFNQNNMIDLITPGGSGSYELPLGVVSG---- 127  
QY 156 GGLGIDNKTKEIKTSVYSLGEYIDLEVFEDNMGMAYSMGLTKFKGNADLIWNSPN 215  
DB 127 -----EDVWVGAGYOYSLDALVDCGRVHTLIGMLDQGMNLSVSVPT 173  
QY 216 YIPNISRILKDDVPNTPLASSKMFKAFRVSKSHSKVKNFIFVVKDLRYLKYDLSY 275  
DB 174 HITQTS-----RYLSAOLHSFVGFRIKTSPEKRVDDYVFPDQKAL---ANHHI 221

QY 276 DSDIDSESV 284  
DB 222 DFTIDGHEL 230

RESULT 5  
ID P96127 PRELIMINARY: PRT: 242 AA.  
AC P96127;  
DT 01-MAY-1997 (Tremblrel. 03, Created)  
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)



=> fil capl; d que 116; fil wpids; d que 124; fil medl; d que 129; fil embase; d que 134; fil caba; d que 139; fil japlo; d que 154; fil biotechds; d que 144; fil litesct; d que 149

Minneapolis 09/004395

FILE 'CAPLUS' ENTERED AT 14:20:44 ON 19 AUG 1999  
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.  
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.  
COPYRIGHT (C) 1999 AMERICAN CHEMICAL SOCIETY (ACS)

Copyright of the articles to which records in this database refer is held by the publishers listed in the PUBLISHER (PB) field (available for records published or updated in Chemical Abstracts after December 26, 1996), unless otherwise indicated in the original publications.

FILE COVERS 1967 - 19 Aug 1999 VOL 131 ISS 8  
FILE LAST UPDATED: 19 Aug 1999 (19990819/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

This file supports Registry for direct browsing and searching of all substance data from the REGISTRY file. Enter HELP FIRST for more information.

L4	186 SEA FILE=CAPLUS ABB=ON	FLAA
L5	1159 SEA FILE=CAPLUS ABB=ON	BORRELIA BURGDOFFERI
L6	215 SEA FILE=CAPLUS ABB=ON	P37
L7	1141 SEA FILE=CAPLUS ABB=ON	37 (W) (KDA OR KILODALTON#)
L8	1541 SEA FILE=CAPLUS ABB=ON	38 (W) (KDA OR KILODALTON#)
L9	192 SEA FILE=CAPLUS ABB=ON	OUTER SHEATH
L10	938 SEA FILE=CAPLUS ABB=ON	LYME
L14	15 SEA FILE=CAPLUS ABB=ON	L5 AND L10 AND (L4 OR L6 OR L7 OR L8
L15	1 SEA FILE=CAPLUS ABB=ON	L14 AND PALIDUM/TI
L16	14 SEA FILE=CAPLUS ABB=ON	L14 NOT L15

FILE LAST UPDATED: 13 AUG 1999  
<19990813/UP>

MOST RECENT DERWENT WEEK 199932  
DERWENT WEEK FOR CHEMICAL CODING: 199932  
DERWENT WEEK FOR POLYMER INDEXING: 199932  
DERWENT WORLD PATENTS INDEX SUBSCRIBER FILE, COVERS 1963 TO DATE

>>> D COST AND SET NOTICE DO NOT REFLECT SUBSCRIBER DISCOUNTS -  
SEE HELP COST <<<

>>> IMPORTANT DERWENT ANNOUNCEMENT ABOUT CHANGES TO CPI  
SUBSCRIBER INDEXING - SEE NEWS <<<

>>> FOR UP-TO-DATE INFORMATION ABOUT ALL 'NEW CONTENT' CHANGES TO  
WPIDS, INCLUDING THE DERWENT CHEMISTRY RESOURCE (DCR),  
PLEASE VISIT <http://www.derwent.com/newcontent.html> <<<

\*\*\*\*\*  
YEAR 2000 FORMAT CHANGES - SEE NEWS  
\*\*\*\*\*





L17 81 SEA FILE=WPIIDS ABB=ON BORRELIA BURGDORFERI  
 L18 178 SEA FILE=WPIIDS ABB=ON LYME  
 L19 13 SEA FILE=WPIIDS ABB=ON P37 OR FLAA  
 L20 33 SEA FILE=WPIIDS ABB=ON (37 OR 38) (2W) (DALTON# OR KDA OR  
 KILODALTON#)  
 L21 2512 SEA FILE=WPIIDS ABB=ON OUTER SHEATH  
 L23 132 SEA FILE=WPIIDS ABB=ON BORRELIA OR BURGDORFERI  
 L24 0 SEA FILE=WPIIDS ABB=ON (L17 OR L18 OR L23) AND (L19 OR L20 OR  
 L21))

FILE 'MEDLINE' ENTERED AT 14:20:47 ON 19 AUG 1999

FILE LAST UPDATED: 16 AUG 1999 (19990816/UP). FILE COVERS 1960 TO DATE.

MEDLINE has been reloaded to reflect the annual MESH changes made by  
 the National Library of Medicine for 1999. Enter HELP RLOAD for details.

OLDMEDLINE, data from 1960 through 1965 from the cumulated Index  
 Medicus (CIM), has been added to MEDLINE. See HELP CONTENT for details.

Left, right, and simultaneous left and right truncation are available in the  
 Basic Index. See HELP SFIELDS for details.

THIS FILE CONTAINS CAS REGISTRY NUMBERS FOR EASY AND ACCURATE  
 SUBSTANCE IDENTIFICATION.

L25 2448 SEA FILE=MEDLINE ABB=ON BORRELIA BURGDORFERI/CT  
 L26 4286 SEA FILE=MEDLINE ABB=ON LYME DISEASE+NT/CT  
 L27 3011 SEA FILE=MEDLINE ABB=ON (37 OR 38) (2W) (DALTON# OR KDA OR  
 KILODALTON#)  
 L28 438 SEA FILE=MEDLINE ABB=ON OUTER SHEATH OR FLAA OR P37  
 L29 15 SEA FILE=MEDLINE ABB=ON (L25 OR L26) AND (L27 OR L28)

FILE 'EMBASE' ENTERED AT 14:20:49 ON 19 AUG 1999

COPYRIGHT (C) 1999 Elsevier Science B.V. All rights reserved.

FILE COVERS 1974 TO 12 AUG 1999 (19990812/ED)

EMBASE has been reloaded. Enter HELP RLOAD for details.

This file contains CAS Registry Numbers for easy and accurate  
 substance identification.

L30 3023 SEA FILE=EMBASE ABB=ON BORRELIA BURGDORFERI/CT  
 L31 3889 SEA FILE=EMBASE ABB=ON LYME DISEASE/CT  
 L32 2505 SEA FILE=EMBASE ABB=ON (37 OR 38) (2W) (DALTON# OR KDA OR  
 KILODALTON#)  
 L33 401 SEA FILE=EMBASE ABB=ON OUTER SHEATH OR FLAA OR P37  
 L34 11 SEA FILE=EMBASE ABB=ON (L30 OR L31) AND (L32 OR L33)

FILE 'CABA' ENTERED AT 14:20:49 ON 19 AUG 1999  
 COPYRIGHT (C) 1999 CAB INTERNATIONAL (CABI)

FILE COVERS 1973 TO 11 AUG 1999 (19990811/ED)

Searched by Barb O'Brien, STIC 308-4740



This file contains CAS Registry Numbers for easy and accurate substance identification.

L35 3846 SEA FILE=CABA ABB=ON BORELIA BURGDOFFERI  
L36 3717 SEA FILE=CABA ABB=ON LYME  
L37 758 SEA FILE=CABA ABB=ON (37 OR 38) (2W) (DALTON# OR KDA OR KILODALTON#)  
L38 133 SEA FILE=CABA ABB=ON OUTER SHEATH OR FLAA OR P37  
L39 13 SEA FILE=CABA ABB=ON (L35 OR L36) AND (L37 OR L38)

FILE 'JAPIO' ENTERED AT 14:20:50 ON 19 AUG 1999  
COPYRIGHT (C) 1999 Japanese Patent Office (JPO) and Japan Patent Information Organization (Japio)

FILE LAST UPDATED: 13 AUG 1999  
<19990813/UP>  
FILE COVERS 1976 TO DATE.

+++++  
YEAR 2000 FORMAT CHANGES - SEE NEWS  
+++++

L50 1 SEA FILE=JAPIO ABB=ON BORELIA BURGDOFFERI  
L51 2 SEA FILE=JAPIO ABB=ON (37 OR 38) (2W) (DALTON# OR KDA OR KILODALTON#)  
L52 283 SEA FILE=JAPIO ABB=ON OUTER SHEATH OR FLAA OR P37  
L53 5 SEA FILE=JAPIO ABB=ON LYME  
L54 0 SEA FILE=JAPIO ABB=ON (L50 OR L53) AND (L51 OR L52)

FILE 'BIOTECHDS' ENTERED AT 14:20:51 ON 19 AUG 1999  
COPYRIGHT (C) 1999 DERWENT INFORMATION LTD

FILE LAST UPDATED: 12 AUG 1999  
<19990812/UP>  
>>> USE OF THIS FILE IS LIMITED TO BIOTECH SUBSCRIBERS <<<  
>>> A THESAURUS IS AVAILABLE IN FIELD CT <<<

L40 119 SEA FILE=BIOTECHDS ABB=ON BORELIA BURGDOFFERI  
L41 128 SEA FILE=BIOTECHDS ABB=ON LYME  
L42 128 SEA FILE=BIOTECHDS ABB=ON (37 OR 38) (2W) (DALTON# OR KDA OR KILODALTON#)  
L43 39 SEA FILE=BIOTECHDS ABB=ON OUTER SHEATH OR FLAA OR P37  
L44 2 SEA FILE=BIOTECHDS ABB=ON (L40 OR L41) AND (L42 OR L43)

FILE 'LIFESCI' ENTERED AT 14:20:52 ON 19 AUG 1999  
COPYRIGHT (C) 1999 Cambridge Scientific Abstracts (CSA)

FILE COVERS 1978 TO 12 AUG 1999 (19990812/ED)

L45 1824 SEA FILE=LIFESCI ABB=ON BORELIA BURGDOFFERI  
L46 1775 SEA FILE=LIFESCI ABB=ON LYME  
L47 1470 SEA FILE=LIFESCI ABB=ON (37 OR 38) (2W) (DALTON# OR KDA OR KILODALTON#)  
L48 273 SEA FILE=LIFESCI ABB=ON OUTER SHEATH OR FLAA OR P37  
L49 11 SEA FILE=LIFESCI ABB=ON (L45 OR L46) AND (L47 OR L48)

=> fill biosis; d que 160

Searched by Barb O'Brien, STIC 308-4740



FILE 'BIOSIS' ENTERED AT 14:23:33 ON 19 AUG 1999  
COPYRIGHT (C) 1999 BIOSIS(R)

FILE COVERS 1969 TO DATE.  
CAS REGISTRY NUMBERS AND CHEMICAL NAMES (CNS) PRESENT  
FROM JANUARY 1969 TO DATE.

RECORDS LAST ADDED: 17 August 1999 (19990817/ED)

The BIOSIS file has been reloaded. Enter HELP RLOAD and HELP REINDEXING  
for details.

L55	4558 SEA FILE=BIOSIS ABB=ON	(B OR BORRELIA) (W) BURGDOFFERI
L56	4454 SEA FILE=BIOSIS ABB=ON	LYME
L57	3361 SEA FILE=BIOSIS ABB=ON	(37 OR 38) (2W) (DALTON# OR KDA OR
		KILODALTON#)
L58	531 SEA FILE=BIOSIS ABB=ON	OUTER SHEATH OR FLAA OR P37
L60	11 SEA FILE=BIOSIS ABB=ON	L55 AND L56 AND (L57 OR L58)

=> dup rem 129,139,160,144,116,149,134

FILE 'MEDLINE' ENTERED AT 14:23:48 ON 19 AUG 1999

FILE 'CABA' ENTERED AT 14:23:48 ON 19 AUG 1999  
COPYRIGHT (C) 1999 CAB INTERNATIONAL (CABI)

FILE 'BIOSIS' ENTERED AT 14:23:48 ON 19 AUG 1999  
COPYRIGHT (C) 1999 BIOSIS(R)

FILE 'BIOTECNDS' ENTERED AT 14:23:48 ON 19 AUG 1999  
COPYRIGHT (C) 1999 DERWENT INFORMATION LTD

FILE 'CAPLUS' ENTERED AT 14:23:48 ON 19 AUG 1999  
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.  
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.  
COPYRIGHT (C) 1999 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'LIFESCI' ENTERED AT 14:23:48 ON 19 AUG 1999  
COPYRIGHT (C) 1999 Cambridge Scientific Abstracts (CSA)

FILE 'EMBASE' ENTERED AT 14:23:48 ON 19 AUG 1999  
COPYRIGHT (C) 1999 Elsevier Science B.V. All rights reserved.

PROCESSING COMPLETED FOR L29  
PROCESSING COMPLETED FOR L39  
PROCESSING COMPLETED FOR L60  
PROCESSING COMPLETED FOR L44  
PROCESSING COMPLETED FOR L16  
PROCESSING COMPLETED FOR L49  
PROCESSING COMPLETED FOR L34

L61 25 DUP REM L29 L39 L60 L44 L16 L49 L34 (52 DUPLICATES REMOVED)

=> d 1b1b ab 161 1-25; fill hom

L61 ANSWER 1 OF 25 CAPLUS COPYRIGHT 1999 ACS  
ACCESSION NUMBER: 1999:451378 CAPLUS  
DOCUMENT NUMBER: 131:83993  
TITLE:

Identification of P37 antigen as  
FLAA, and the use of recombinant P37  
as a diagnostic reagent for Lyme disease  
Searched by Barb O'Brien, STIC 308-4740



INVENTOR(S): Gilmore, Robert D., Jr.; Johnson, Barbara J. B.  
 PATENT ASSIGNEE(S): Biometreux, Inc., USA  
 SOURCE: PCT Int. Appl., 34 pp.  
 CODEN: PIXXD2  
 Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 1  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9935272	A1	19990715	WO 1999-US196	19990106
-----				
AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GU, HN, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, AM, AZ, BY, BG, CH, CY, CG, CI, CM, GA, GN, GW, ML, MR, NE, NL, PT, SE, BF, BJ, CF, CG, CI, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG	PRIORITY APPL. INFO.: US 1998-4395 19980108			

AB The invention provides reagents and improved methods for serodiagnostics of Lyme disease, particularly in persons with recently acquired infection. The invention describes the definitive identification of the antigen P37 as FlaA, an outer sheath protein of the periplasmic flagella of *Borrelia burgdorferi*. The invention also demonstrates that P37 (FlaA) is a prominent antigen in the early humoral immune response to B. burgdorferi infection and is significantly suitable for use in improved serol. tests for exposure to Lyme disease spirochetes. The invention provides an assay for detecting Lyme disease infection comprising obtaining a serum sample from a patient, contacting said sample with recombinant P37 (FlaA), and detecting any antibody specifically bound to said protein. In a preferred embodiment, the recombinant P37 (FlaA) antigen is produced as a fusion protein with the T7 gene 10 product.

DUPLICATE 1

L61 ANSWER 2 OF 25 MEDLINE  
 ACCESSION NUMBER: 1999143208  
 DOCUMENT NUMBER: 99143208

TITLE: The *Borrelia burgdorferi* 37-kilodalton

AUTHOR: Gilmore R D Jr; Murphree R L; James A M; Sullivan S A; Johnson B J  
 CORPORATE SOURCE: Division of Vector-Borne Infectious Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Public Health Service, U.S. Department of Health and Human Services, Fort Collins, Colorado, USA..

SOURCE: JOURNAL OF CLINICAL MICROBIOLOGY, (1999 Mar) 37 (3) 548-52.  
 Journal code: HSH. ISSN: 0095-1137.

PUB. COUNTRY: United States  
 LANGUAGE: English  
 FILE SEGMENT: Priority Journals

ENTRY MONTH: 199905  
 ENTRY WEEK: 19990503

AB The 37-kDa protein (P37) of *Borrelia*

*burgdorferi* is an antigen that elicits an early immunoglobulin M (IgM) antibody response in Lyme disease patients. The P37 gene was cloned from a B. burgdorferi genomic library by screening with antibody from a Lyme disease patient who had developed a prominent humoral response

Searched by Barb O'Brien, STIC 308-4740





to the P37 antigen. DNA sequence analysis of this clone revealed the identity of P37 to be Fla, an outer sheath protein of the periplasmic flagella. Recombinant P37 expression was accomplished in *Escherichia coli* by using a gene construct with the leader peptide deleted and fused to a 38 kDa E. coli protein. The recombinant antigen was reactive in IgM immunoblots using serum samples from patients clinically diagnosed with early Lyme disease that had been scored positive for B. burgdorferi anti-P37 reactivity. Lyme disease patient samples serologically negative for the B. burgdorferi P37 protein did not react with the recombinant. Recombinant P37 may be a useful component of a set of defined antigens for the serodiagnosis of early Lyme disease. This protein can be utilized as a marker in diagnostic immunoblots, aiding in the standardization of the present generation of IgM serologic tests.

DUPLICATE 2

L61 ANSWER 3 OF 25 MEDLINE

MEDLINE

ACCESSION NUMBER: 1999081739

DOCUMENT NUMBER: 99081739

TITLE:

The immunoglobulin (IgG) antibody response to OspA and OspB correlates with severe and prolonged Lyme arthritis and the IgG response to P35 correlates with mild and brief arthritis.

AUTHOR:

Akin E; McHugh G L; Flavell R A; Fikrig E; Steere A C

CORPORATE SOURCE:

School of Medicine, New England Medical Center, Tufts University

CONTRACT NUMBER:

AR-20538 (NIAMS)

SOURCE:

INFECTION AND IMMUNITY, (1999 Jan) 67 (1) 173-81.

PUB. COUNTRY:

United States

LANGUAGE:

English

FILE SEGMENT:

Priority Journals; Cancer Journals

ENTRY MONTH:

199903

ENTRY WEEK:

19990305

AB

In an effort to implicate immune responses to specific Borrelia

burgdorferi proteins that may have a role in chronic Lyme arthritis, we studied the natural history of the antibody response to B. burgdorferi in serial serum samples from 25 patients monitored throughout the course of Lyme disease. In these patients, the immunoglobulin G (IgM) and IgG antibody responses to 10 recombinant B. burgdorferi proteins, determined during early infection, early arthritis, and maximal arthritis, were correlated with the severity and duration of maximal arthritis. The earliest responses were usually to outer surface protein C (OspC), P35, P37, and P41; reactivity with OspE, OspF, P39, and P93 often developed weeks later; and months to years later, 64% of patients had responses to OspA and OspB. During early infection and early arthritis, the levels of IgG antibody to P35 correlated inversely with the subsequent severity or duration of maximal arthritis. In contrast, during periods of maximal arthritis, the levels of IgG antibody to OspA and OspB, especially to a C-terminal epitope of OspA, correlated directly with the severity and duration of arthritis. Thus, the higher the IgG antibody response to P35 earlier in the infection, the milder and briefer the subsequent arthritis, whereas during maximal arthritis, the higher the IgG response to OspA and OspB, the more severe and prolonged the arthritis.

DUPLICATE 3

L61 ANSWER 4 OF 25 MEDLINE

MEDLINE

ACCESSION NUMBER: 1998233738

DOCUMENT NUMBER: 98233738

TITLE:

Structure and expression of the Fla periplasmic

flagellar protein of Borrelia burgdorferi.

AUTHOR:

Ge Y; Li C; Corum L; Slaughter C A; Charon N W

CORPORATE SOURCE:

Department of Microbiology and Immunology, Robert C. Byrd

Searched by Barb O'Brien, STIC 308-4740



Health Sciences Center, West Virginia University,  
Morgantown 26506-9177, USA.  
CONTRACT NUMBER: AI29743 (NIAID)  
JOURNAL OF BACTERIOLOGY, (1998 May) 180 (9) 2418-25.  
JOURNAL CODE: HH3. ISSN: 0021-9193.  
PUB. COUNTRY: United States  
LANGUAGE: English  
FILE SEGMENT: Priority Journals  
ENTRY MONTH: 199808  
ENTRY WEEK: 19980801  
AB The spirochete which causes Lyme disease, *Borrelia burgdorferi*, has many features common to other spirochete species. Outermost is a membrane sheath, and within this sheath are the cell cylinder and periplasmic flagella (PFs). The PFs are subterminally attached to the cell cylinder and overlap in the center of the cell. Most descriptions of the *B. burgdorferi* flagellar filaments indicate that these organelles consist of only one flagellin protein (FlaB). In contrast, the PFs from other spirochete species are comprised of an outer layer of FlaA and a core of FlaB. We recently found that a FlaA homolog was expressed in *B. burgdorferi* and that it mapped in a flaA che operon. These results led us to analyze the PFs and FlaA of *B. burgdorferi* in detail. Using Triton X-100 to remove the outer membrane and isolate the PFs, we found that the 38.0-kDa FlaA protein purified with the PFs in association with the 41.0-kDa FlaB protein. On the other hand, purifying the PFs by using Sarkosyl resulted in no FlaA in the isolated PFs. Sarkosyl has been used by others to purify *B. burgdorferi* PFs, and our results explain in part their failure to find FlaA. Unlike other spirochetes, *B. burgdorferi* FlaA was expressed at a lower level than FlaB. In characterizing FlaA, we found that it was posttranslationally modified by glycosylation, and thus it resembles its counterpart from *Serpulina hyalogenes*. We also tested if FlaA was synthesized in a spontaneously occurring PF mutant of *B. burgdorferi* (HB19Fla-). Although this mutant still synthesized FlaA message in amounts similar to the wild-type amounts, it failed to synthesize FlaA protein. These results suggest that, in agreement with data found for FlaB and other spirochete flagellar proteins, FlaA is likely to be regulated on the translational level. Western blot analysis using *Treponema pallidum* anti-FlaA serum indicated that FlaA was antigenically well conserved in several spirochete species. Taken together, the results indicate that both FlaA and FlaB comprise the PFs of *B. burgdorferi* and that they are regulated differently from flagellin proteins of other bacteria.

L61 ANSWER 5 OF 25 CABA COPYRIGHT 1999 CABI  
ACCESSION NUMBER: 1998:163535 CABA  
DOCUMENT NUMBER: 980505876  
TITLE: Cloning and expression of the 44-kilodalton major outer membrane protein gene of the human granulocytic ehrlichiosis agent and application of the recombinant protein to serodiagnostics  
AUTHOR: Zhi, N.; Ohashi, N.; Rikihisa, Y.; Horowitz, H. W.; Wormser, G. P.; Hechemy, K.  
CORPORATE SOURCE: Department of Veterinary Biosciences, College of Veterinary Medicine, The Ohio State University, Columbus, OH, USA.  
SOURCE: Journal of Clinical Microbiology, (1998) Vol. 36, No. 6, pp. 1666-1673. 31 ref.  
ISSN: 0095-1137  
DOCUMENT TYPE: Journal  
LANGUAGE: English  
AB A 44-kDa major outer membrane protein of the human granulocytic

Searched by Barb O'Brien, STIC 308-4740



ehrlichiosis (HGE) agent is an immunodominant antigen in human infection. A gene encoding this protein was cloned and sequenced. Southern blot results revealed the existence of multigenes homologous to the p44 gene in the genome of the HGE agent. The recombinant 44-kDa protein (rp44) was expressed by using expression vector pET30a. The reactivity of the affinity-purified rp44 was evaluated by Western immunoblot analysis and dot blot immunoassay. Western immunoblot analysis showed that mouse anti-rp44 serum reacted with 44- to 42-kDa proteins in 6 different HGE agent strains tested except strain 2, in which 3 proteins of 42, 40, and 38 kDa were recognized. 11 HGE patient serum samples (from patients in New York State, USA), a horse anti-HGE serum, and a horse anti-Ehrlichia equi serum recognized the rp44 protein. This suggests that rp44 is an HGE-E. equi group-specific antigen. Neither human anti-E. chaffeensis serum nor rabbit anti-Borrelia burgdorferi serum reacted with the HGE agent and B. burgdorferi reacted positively with rp44 and the HGE agent. Sera from 20 HGE patients with indirect fluorescent reactions in a dot ranging from 1:20 to 1:2560 gave distinct positive reactions in a dot immunoblot assay. There was a positive correlation between the colour densities of the dot reactions and the IFA titres when >50 ng of recombinant antigen per dot was used. The use of the affinity-purified rp44 protein as antigen would provide a more specific, consistent, and simpler serodiagnostics for HGE than the use of whole infected cells or purified HGE agents.

DUPLICATE 5

L61 ANSWER 6 OF 25 MEDLINE  
ACCESSION NUMBER: 1999023007  
DOCUMENT NUMBER: 99023007  
TITLE: Differential expression of Borrelia burgdorferi genes during erythema migrans and Lyme arthritis.

AUTHOR: Fikrig E, Feng W, Aversa J, Schoen R T, Flavell R A  
CORPORATE SOURCE: Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut 06520-8031, USA.  
SOURCE: JOURNAL OF INFECTIOUS DISEASES, (1998 Oct) 178 (4) 1198-201.

PUB. COUNTRY: United States  
JOURNAL CODE: IH3. ISSN: 0022-1899.  
LANGUAGE: English  
FILE SEGMENT: Abridged Index Medicus Journals; Priority Journals  
ENTRY MONTH: 199901  
ENTRY WEEK: 19990104

AB Borrelia burgdorferi, the agent of Lyme disease, selectively expresses genes in the arthropod vector and mammalian host. Specific B. burgdorferi gene expression during human infection was examined in tissue specimens, using RNA-polymerase chain reaction, from 3 patients with Lyme disease. ospA was investigated because ospA is down-regulated by B. burgdorferi in ticks during engorgement and is a vaccine candidate in phase III clinical trials. p35 and p37 were also assessed because these genes are induced by spirochetes during murine Lyme borreliosis and play roles in protective immunity. p35 and p37 mRNA were detected in erythema migrans biopsy specimens from 2 patients and in the synovium of 1 patient with Lyme arthritis. ospA mRNA was not identified in any of these tissues. These data show that ospA is repressed while p35 and p37 are induced in human infection; these results are the first direct demonstration of differential B. burgdorferi gene expression during Lyme disease.

L61 ANSWER 7 OF 25 CAPLUS COPYRIGHT 1999 ACS  
ACCESSION NUMBER: 1999:50456 CAPLUS  
DOCUMENT NUMBER: 130:94463  
TITLE: Borrelia burgdorferi p35 and

p37 proteins, expressed in vivo, elicit

Searched by Barb O'Brien, STIC 308-4740



protective immunity. [Erratum to document cited in CA127:107928]  
 AUTHOR(S): Fikrig, Erol; Barthold, Stephen W.; Sun, Wei; Feng, Wen; Telford, Sam R., III; Flavell, Richard A. Section Rheumatology, Dep. Internal Med., Yale Univ. Sch. Med., New Haven, CT, 06520, USA  
 SOURCE: Immunity (1998), 9(5), No pp. given  
 CODEN: IUNIEH; ISSN: 1074-7613  
 PUBLISHER: Cell Press  
 DOCUMENT TYPE: Journal  
 LANGUAGE: English  
 AB The initial portion of the p35 sequence reported in this paper is not a *Borrelia burgdorferi* sequence. The first 114 nucleotides of the p35 sequence are part of the pBluescript vector sequence. The remaining 813 nucleotides of the gene are correct. The complete gene sequence (see Fraser C. M., et al. [1997]. Genomic sequence of a *Lyme* disease spirochaete, *Borrelia burgdorferi*. Nature 390, 580-586) is designated as B. burgdorferi B31 open reading frame BBK32.  
 L61 ANSWER 8 OF 25 CAPLUS COPYRIGHT 1999 ACS  
 ACCESSION NUMBER: 1997:746160 CAPLUS  
 DOCUMENT NUMBER: 128:31107  
 TITLE: *Borrelia burgdorferi* proteins expressed during infection and methods and compositions for prevention, treatment, and diagnosis of *Lyme* disease  
 INVENTOR(S): Fikrig, Erol; Suk, Kyoungho; Barthold, Stephen W.; Flavell, Richard A.  
 PATENT ASSIGNEE(S): Yale University, USA; Fikrig, Erol; Suk, Kyoungho; Barthold, Stephen W.; Flavell, Richard A.  
 SOURCE: PCT Int. Appl., 115 pp.  
 CODEN: PIXXD2  
 Patent  
 English  
 FAMILY ACC. NUM. COUNT: 1  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9742325	A1	19971113	WO 1996-US6610	19960508
W: CA, JP, US				
RW: AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE				
EP 915977	A1	19990519	EP 1996-915650	19960508
R: AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				
PRIORITY APPLN. INFO.:			WO 1996-US6610	19960508

AB Methods and comps. for the prevention, treatment and diagnosis of *Lyme* disease are disclosed. Novel B. burgdorferi proteins which are produced during infection of a host but are not expressed by in vitro cultures of B. burgdorferi, as well as genes for these proteins, are identified. Vaccines comprising the novel B. burgdorferi proteins, antibodies to these proteins, and diagnostic use of the proteins and antibodies are claimed. A method for identifying bacterial genes that are selectively expressed in vivo is also disclosed. Genes expressed during infection of mice but not expressed by in vitro cultures of B. burgdorferi were identified by differential screening of a B. burgdorferi DNA library with antisera from mice inoculated with killed B. burgdorferi and with antisera from mice infected with live B. burgdorferi spirochetes. Several genes were isolated and sequenced. The genes for the P21, P35 and P37 proteins were mapped to sep. plasmids. Antibodies to these proteins were prepd. A humoral response to P35 and P37 was detected in humans infected with B. burgdorferi. S.C. injection of mice

Searched by Barb O'Brien, STIC 308-4740





with P35 and P37 may protect mice from challenge with B. burgdorferi, but P21 did not protect mice from tick-mediated infection.

DUPLICATE 6

Medline

L61 ANSWER 9 OF 25 MEDLINE  
 97342782  
 97342782  
 DOCUMENT NUMBER: 97342782  
 TITLE: FlaA, a putative flagellar outer sheath protein, is not an immunodominant antigen associated with Lyme disease.

AUTHOR: Ge Y; Charon N W  
 DEPARTMENT OF MICROBIOLOGY AND IMMUNOLOGY, WEST VIRGINIA UNIVERSITY, Morgantown 26506-9177, USA.  
 CONTRACT NUMBER: AI29743 (NIAID)  
 SOURCE: INFECTION AND IMMUNITY, (1997 Jul) 65 (7) 2992-5.  
 PUB. COUNTRY: United States  
 LANGUAGE: English  
 FILE SEGMENT: Priority Journals; Cancer Journals  
 ENTRY MONTH: 199709  
 ENTRY WEEK: 19970904  
 AB **FlaA** was recently found to be associated with flagellar filaments of *Borrelia burgdorferi*. We tested whether antibodies to this protein are a good indicator of infection, as antibodies to **FlaA** proteins in other spirochetal infections show an increase in titer. Although overproduction of intact **FlaA** was highly toxic to *Escherichia coli*, truncated proteins which lacked the N-terminal signal sequence could be successfully overexpressed. Immunoblotting with sera from mammalian hosts infected with *B. burgdorferi* indicated that **FlaA** is not an immunodominant antigen in Lyme disease. However, sera from two patients reacted with both recombinant and native **FlaA** protein, suggesting that *B. burgdorferi FlaA* was antigenic and expressed in vivo.

L61 ANSWER 10 OF 25 MEDLINE  
 97312006  
 97312006  
 MEDLINE  
 ACCESSION NUMBER: 97312006  
 DOCUMENT NUMBER: 97312006  
 TITLE: The flag motility operon of *Borrelia burgdorferi* is initiated by a sigma 70-like promoter.  
 AUTHOR: Ge Y; Old I G; Charon N W  
 DEPARTMENT OF MICROBIOLOGY, West Virginia University, Morgantown 26506-9177, USA.  
 CONTRACT NUMBER: AI2743 (NIAID)  
 SOURCE: MICROBIOLOGY, (1997 May) 143 ( Pt 5) 1681-90.  
 PUB. COUNTRY: ENGLAND: United Kingdom  
 LANGUAGE: English  
 FILE SEGMENT: Priority Journals  
 OTHER SOURCE: GENBANK-U62901; GENBANK-U66699  
 ENTRY MONTH: 199709  
 AB A cluster of flagellar genes of *Borrelia burgdorferi* was identified and sequenced. This cluster comprises an operon, designated the flag operon, which is initiated by a sigma 70-like promoter. The flag operon consists of flbF (function unknown), flgK (encoding HAP1), flgL (encoding HAP3) and flaA (function unknown), and maps at 185 kb on the chromosome. In other bacteria, the hook-associated proteins HAP1 and HAP3 connect the flagellar filament to the hook and are required for the last stage of flagellar assembly. Reverse transcriptase-PCR analysis indicated that flbF through to flbF are transcribed as a single mRNA, and primer extension analysis revealed that transcription of the flag operon is initiated by a sigma 70-like promoter upstream of flbF. Subcloning the flag promoter element into a promoter probe cat vector revealed that the flag promoter element

Searched by Barb O'Brien, STIC 308-4740



had strong activity in both *Escherichia coli* and *Salmonella typhimurium*. In addition, when this construct was transformed into a *flaA* mutant of *S. typhimurium* which lacked a functional flagellar-specific sigma 28 factor, the *flg* promoter was still functional. Based on these results, the promoter element of the flagellin gene (*fla*), hereafter referred to as (fla) was re-examined. flab encodes the flagellar filament protein, and a sigma gp33-34-like promoter has been reported to be involved in the transcription of this gene. A transcriptional start point was found 1 bp downstream of the reported start site. The sequence around -10 and -35 are consistent with the presence of a sigma 70-like promoter in addition to the putative sigma gp33-34-like promoter for flab. In contrast to the *flg* promoter element, no activity was detected after subcloning a flab promoter element into the promoter probe cat vector. Because a sigma 70-like promoter rather than a unique flagellar sigma factor is involved in the later stage of flagellar assembly, the regulation of *B. burgdorferi* flagellar genes is evidently different from that of other bacteria.

DUPLICATE 7

L61 ANSWER 11 OF 25 MEDLINE

ACCESSION NUMBER: 97144545 MEDLINE

DOCUMENT NUMBER: 97144545

TITLE:

An unexpected *flaA* homolog is present and expressed in *Borrelia burgdorferi*.

AUTHOR:

Ge Y; Charon N W

CORPORATE SOURCE:

Department of Microbiology and Immunology, West Virginia University, Morgantown 26506-9177, USA.

CONTRACT NUMBER:

A129743 (NIAID)

SOURCE:

JOURNAL OF BACTERIOLOGY, (1997 Jan) 179 (2) 552-6. Journal code: HH3. ISSN: 0021-9193.

PUB. COUNTRY:

United States

LANGUAGE:

English

FILE SEGMENT:

Priority Journals

OTHER SOURCE:

GENBANK-U62900

ENTRY MONTH:

199704

AB

Most investigators have assumed that the periplasmic flagella (PFs) of *Borrelia burgdorferi* are composed of only one flagellin protein. The PFs of most other spirochete species are complex: these PFs contain an outer sheath of flab proteins and a core filament of flab proteins. During an analysis of a chemotaxis gene cluster of *B. burgdorferi* 212, we were surprised to find a *flaA* gene homolog with a deduced polypeptide having 54 to 58% similarity to flab from other spirochetes. Like other flab proteins, *B. burgdorferi* flab has a conserved signal sequence at its N terminus. Based on reverse transcription-PCR and primer extension analysis, this flab homolog and five chemotaxis genes constitute a motility-chemotaxis operon. Immunoblots using anti-flab serum from *Treponema pallidum* and a lysate of *B. burgdorferi* showed strong reactivity to a protein of 38.0 kDa, which is consistent with the expression of flab in growing cells.

DUPLICATE 8

L61 ANSWER 12 OF 25 MEDLINE

ACCESSION NUMBER: 97318894 MEDLINE

DOCUMENT NUMBER: 97318894

TITLE:

*Borrelia burgdorferi* P35 and P37 proteins, expressed in vivo, elicit protective immunity [published erratum appears in Immunity 1998 Nov;9(5):following 755].

AUTHOR:

Flakrig E; Barthold S W; Sun W; Feng W; Telford S R 3rd; Flavell R A

CORPORATE SOURCE:

Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut 06520, USA.

CONTRACT NUMBER:

A1-26815 (NIAID)

A1-30548 (NIAID)

Searched by Barb O'Brien, STIC 308-4740



AR-40452 (NIAMS)

IMMUNITY, (1997 May) 6 (5) 531-9.

Journal code: CCF. ISSN: 1074-7613.

United States

Journal; Article; (JOURNAL ARTICLE)

English

Priority Journals

199708

ENTRY MONTH:

FILE SEGMENT:

LANGUAGE:

PUB. COUNTRY:

SOURCE:

p35 and p37 are Borrelia burgdorferi genes encoding 35 and

37 kDa proteins. The gene products were identified by

differential screening of a B. burgdorferi expression library with sera

from B. burgdorferi infected- and B. burgdorferi-hyperimmunized mice.

Northern blot and RT-PCR analyses confirmed that these genes were

selectively expressed in vivo. ELISA, using p35 and p37, showed

that infected mice (5 of 5, 100%) and patients (31 of 43, 72%) with Lyme

borreliosis developed p35 or p37 antibodies. Mice developed peak

IgG titers to p35 and p37 within 30 days, followed by decline.

Mice given both p35 and p37 antisera were protected from

challenge with 10(2) B. burgdorferi, and p35 and p37 antisera

also afforded protection when administered 24 hr after spirochete

challenge. The use of in vivo-expressed antigens such as p35 and

p37 represents a new approach for Lyme disease serodiagnostics and

for understanding the role of B. burgdorferi-specific immune responses in

host immunity.

L61 ANSWER 13 OF 25 MEDLINE

ACCESSION NUMBER: 97417814 MEDLINE

DOCUMENT NUMBER: 97417814

TITLE:

Molecular characterization of a flagellar/chemotaxis operon

in the spirochete Borrelia burgdorferi.

AUTHOR:

Ge Y; Charon N W

CORPORATE SOURCE:

Department of Microbiology and Immunology, West Virginia

University, Robert C. Byrd Health Sciences Center,

Morgantown 26506-9177, USA.

CONTRACT NUMBER:

A129743 (NIAID)

SOURCE:

FEMS MICROBIOLOGY LETTERS, (1997 Aug 15) 153 (2) 425-31.

PUB. COUNTRY:

Netherlands

Journal; Article; (JOURNAL ARTICLE)

English

FILE SEGMENT:

Priority Journals

OTHER SOURCE:

GENBANK-U61498

ENTRY MONTH:

199711

AB A chemotaxis gene cluster from Borrelia burgdorferi, the spirochete that

causes Lyme disease, was cloned, sequenced, and analyzed. This cluster

contained three chemotaxis gene homologs (cheA, cheW and cheY) and an open

reading frame we identified as cheX. Although the major functional domains

for B. burgdorferi cheW and cheY were well conserved, the size of cheW was

significantly different from the homolog of other bacteria. Phylogenetic

analysis of cheY indicated that B. burgdorferi constitutes a distinct

branch with Treponema pallidum and is closely associated with Archaea and

Gram-positive bacteria. RT-PCR analysis indicated that the chemotaxis

genes and the upstream flagellar gene flhA constitute an operon.

Western blot analysis using antibody to Escherichia coli cheA resulted in

two reactive proteins in the cell lysates of B. burgdorferi that is

consistent with two cheA homologs being present in this organism. The

results taken together suggest both similarities and differences in the

chemotaxis apparatus of B. burgdorferi compared to those of other

bacteria.

L61 ANSWER 14 OF 25 CAPLUS COPYRIGHT 1999 ACS

ACCESSION NUMBER: 1997:298657 CAPLUS

Searched by Barb O'Brien, STIC 308-4740



DOCUMENT NUMBER: 126:339490  
 TITLE: A cheA chew operon in *Borrelia burgdorferi*, the agent of Lyme disease  
 AUTHOR(S): Trueba, G. A.; Old, I. G.; Saint Glrons, I.; Johnson, R. C.  
 CORPORATE SOURCE: Department of Microbiology, University of Minnesota Medical School, Minneapolis, MN, 55455, USA  
 SOURCE: Res. Microbiol. (1997), 148(3), 191-200  
 CODEN: RMCREW; ISSN: 0923-2508  
 PUBLISHER: Elsevier  
 DOCUMENT TYPE: Journal  
 LANGUAGE: English  
 AB *Borrelia burgdorferi* sensu stricto homologs of cheA and cheW were cloned and characterized. A combination of strategies such as polymerase chain reaction (PCR) using degenerate primers, random-primed gene walking PCR and construction of a lambda library were used to identify the putative cheA gene. Sequence anal. of the DNA fragments obtained from the CT strain identified a 2,592-bp open reading frame (ORF) encoding an 864-amino-acid protein with significant similarity (53-64.6% identical residues) to the cheA of several genera of eubacteria. In particular, hallmarks of a histidine kinase family were found such as the location of the histidine autophosphorylation domain very close to the N terminus and the nucleotide-binding site. A second ORF located immediately downstream from the putative borrelial cheA gene encoded a 195-amino-acid protein which displayed a high level of similarity to bacterial CheW. Using reverse transcription PCR, we demonstrated that cheA and cheW form an operon with an upstream, unidentified ORF. The cheA and cheW homologs were located at 722-737 kbp, 738-768 kbp and 743-824 kbp on the linear chromosomes of *B. burgdorferi* sensu stricto, *B. garinii* and *B. atzei*, resp. Identification of cheA and cheW is the first step toward elucidation of a possible role of chemotaxis in virulence of the Lyme disease borreliae.

L61 ANSWER 15 OF 25 CAPLUS COPYRIGHT 1999 ACS  
 ACCESSION NUMBER: 1997:729 CAPLUS  
 DOCUMENT NUMBER: 126:103100  
 TITLE: Recombinant mycobacteria expressing Borrelia antigen on their surface and use of recombinant mycobacteria as vaccines for Lyme disease  
 INVENTOR(S): Stover, Charles K.  
 PATENT ASSIGNEE(S): MedImmune, Inc., USA  
 SOURCE: U.S., 112 pp. Cont.-in-part of U.S. Ser. No. 780,261, abandoned.  
 CODEN: USXXAM  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 2  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 5583038	A	19961210	US 1992-977630	19921117
PRIORITY APPL. INFO.:			US 1991-780261	19911021

AB An expression vector for expressing a protein in mycobacteria, which comprises a first DNA sequence encoding at least a secretion signal of a lipoprotein, and a second DNA sequence encoding a protein, or fragment thereof, which is heterologous to the mycobacteria which express the protein or fragment as described. The mycobacteria express a fusion protein comprising a lipoprotein or lipoprotein segment and the protein or fragment. Such expression vectors increase the immunogenicity of the protein or fragment by enabling the protein or fragment to be expressed on the surface of the mycobacteria. Mycobacteria which may be transformed

Searched by Barb O'Brien, STIC 308-4740





with the expression vector include mycobacteria such as BCG. The formation of live bacterial vaccines against *Lyme* disease wherein the mycobacteria express a surface protein of *Borrelia burgdorferi*, the causative agent of *Lyme* disease. Recombinant BCG expressing a chimeric gene comprising M. tuberculosis 19 kDa antigen promoter and signal sequence fused to B. burgdorferi OspA antigen gene were prepd. The OspA antigen was presented on the surface of BCG. Mice vaccinated with these recombinant BCG were protected from challenge with B. burgdorferi.

L61 ANSWER 16 OF 25 MEDLINE

ACCESSION NUMBER: 96110942 MEDLINE

DOCUMENT NUMBER: 96110942

TITLE: A glyceraldehyde-3-phosphate dehydrogenase homolog in

AUTHOR:

Anda P; Gebbia J A; Backenson P B; Coleman J L; Benach J L  
Centro Nacional de Microbiología, Virología e Inmunología  
Sanitarias, Instituto de Salud Carlos III, Madrid, Spain.

CONTRACT NUMBER:

AI-27044 (NIAID)  
AR-40455 (NIAMS)

SOURCE:

INFECTION AND IMMUNITY, (1996 Jan) 64 (1) 262-8.  
Journal code: G07, ISSN: 0019-9567.

PUB. COUNTRY:

United States  
Journal; Article; (JOURNAL ARTICLE)

LANGUAGE:

English  
Priority Journals; Cancer Journals

OTHER SOURCE:

GENBANK-U28760; GENBANK-U28761  
199605

AB A polyclonal antibody recognized a 38.5-kDa

polypeptide with amino-terminal sequence identity to conserved regions of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in *Borrelia burgdorferi*, the *Lyme* disease agent, and *Borrelia hermsii*, an agent of American relapsing fever. This monoclonal antibody also recognized GAPDH from other pathogenic spirochetes and other prokaryotes as well. GAPDH activity was detected in sonicates of both B. burgdorferi and B. hermsii but not in live, intact organisms, indicating the possibility of a subsurface localization for the *Borrelia* GAPDH activity. Degenerate primers constructed from highly conserved regions of gapdh of other prokaryotes successfully amplified this gene homolog in both B. burgdorferi and B. hermsii. Nuclei acid and deduced amino acid sequence analysis of the 838-bp probes for each *Borrelia* indicated 93.9% identity between B. burgdorferi and B. hermsii at the amino acid level. Amino acid identities of B. burgdorferi and B. hermsii with *Bacillus stearothermophilus* were 59.2% and 58.8% respectively. Southern hybridization studies indicated that the gene encoding GAPDH is located on the chromosome of each *Borrelia*. In other bacterial species, GAPDH has other functions in addition to its traditional enzymatic role in the glycolytic pathway. GAPDH may play a similar role in *Borrelia*.

L61 ANSWER 17 OF 25 MEDLINE

ACCESSION NUMBER: 96167817 MEDLINE

DOCUMENT NUMBER: 96167817

TITLE:

Expression and sequence analysis of a *Treponema pallidum* gene, *tpn38(b)*, encoding an exported protein with homology to T. pallidum and *Borrelia burgdorferi* proteins.

AUTHOR:

Stamm L V; Hardham J M; Frye J G  
Department of Epidemiology, School of Public Health,  
University of North Carolina, Chapel Hill 27599-7400, USA..

CORPORATE SOURCE:

CONTRACT NUMBER: 1 U01 AI31496 (NIAID)  
3 T32 AI07001 (NIAID)

AI24976 (NIAID)

Searched by Barb O'Brien, STIC 308-4740



SOURCE: FEMS MICROBIOLOGY LETTERS, (1996 Jan 1) 135 (1) 57-63.  
 PUB. COUNTRY: Netherlands  
 Journal code: FML. ISSN: 0378-1097.  
 LANGUAGE: English  
 FILE SEGMENT: Priorly Journals  
 OTHER SOURCE: GENBANK-U12861  
 ENTRY MONTH: 199606  
 AB An Escherichia coli clone containing recombinant plasmid C19 was identified from a Treponema pallidum genomic DNA library by in situ immunosassay. E. coli maxicells containing pC19 synthesized a treponemal protein doublet of 39.2 and 38.2 kDa, designated TPN38(b). Pulse-chase and protein processing studies showed that TPN38(b) is synthesized with a cleavable amino-terminal signal peptide. A 2.0-kb fragment of pC19 containing the tpn38(b) gene was subcloned and sequenced. The tpn38(b) gene is 1029 nucleotides long and encodes a protein of 343 amino acids with a calculated molecular mass of 37.9 kDa. The deduced amino acid sequence of TPN38(b) has homology with the T. pallidum TPN35 lipoprotein and the Borrelia burgdorferi BmpA, BmpB, BmpC, and BmpD proteins.

L61 ANSWER 18 OF 25 MEDLINE  
 MEDLINE  
 ACCESSION NUMBER: 96352974  
 DOCUMENT NUMBER: 96352974  
 TITLE: Evolution of the serologic response to Borrelia burgdorferi in treated patients with culture-confirmed erythema migrans.  
 AUTHOR: Agüero-Rosenfeld M E; Nowakowski J; Bittker S; Cooper D; Nadelman R B; Wormser G P  
 CORPORATE SOURCE: Department of Pathology, New York Medical College, Valhalla, USA.  
 CONTRACT NUMBER: U50-CCU 210280-01 (NIAMS)  
 R01-AR41508 (NIAMS)  
 R01-AR43135  
 SOURCE: JOURNAL OF CLINICAL MICROBIOLOGY, (1996 Jan) 34 (1) 1-9.  
 Journal code: HSH. ISSN: 0095-1137.  
 PUB. COUNTRY: United States  
 LANGUAGE: English  
 FILE SEGMENT: Priorly Journals  
 ENTRY MONTH: 199612  
 AB We investigated the appearance and evolution of immunoglobulin M (IgM) and IgG antibodies to Borrelia burgdorferi in 46 patients with culture-proven erythema migrans (EM). All patients received antimicrobial treatment and were prospectively evaluated for up to 1 year. A total of 257 serially collected serum samples were tested by commercial IgG-IgM enzyme-linked immunosorbent assay and separate IgM and IgG immunoblots (IBs). At the baseline, 33% of the patients had a positive ELISA result and 43% of the patients had a positive IgM IB result by using the criteria of the Centers for Disease Control and Prevention-Association of State and Territorial Public Health Laboratory Directors for the interpretation of IB results. Positive serology at the baseline and the rate of seroconversion correlated directly with disease duration and/or evidence of dissemination prior to treatment. At days 8 to 14 after the baseline, 91% of patients had a positive ELISA result and/or IgM IB result. Peak IgM antibody levels were seen at this time in patients with localized or disseminated disease. The most frequent IgM bands at the baseline and the peak were of 24 kDa (OspC), 41 kDa, and 37 kDa. Although 89% of the patients developed IgG antibodies as determined at a follow-up examination, only 22% were positive by the IgG IB criteria of the Centers for Disease Control and Prevention-Association of State and Territorial Public Health Laboratory Directors. The persistence of antibodies was directly related to disease duration and/or dissemination prior to

Searched by Barb O'Brien, STIC 308-4740

DUPLICATE 11



treatment. Since IgM antibodies to the 24- and 41-kDa antigens remained detectable for long periods, 38% of IgM IBS were still positive at 1 year postbaseline. IgM to antigens of 39, 58, 60, 66, or 93 kDa, conversely, were most often seen in sera obtained within 1 month postbaseline. Their presence may be of assistance in confirming a recent infection with B. burgdorferi in individuals living in areas where Lyme disease is endemic.

L61 ANSWER 19 OF 25 BIOTECCHDS COPYRIGHT 1999 DERWENT INFORMATION LTD

TITLE:

New Borrelia burgdorferi polypeptide;

associated with highly infective strains used for

detection of Lyme disease and in vaccine

AUTHOR:

Norris S J; Barbour A G

PATENT ASSIGNEE:

Univ. Texas-Syst.

PATENT INFO:

WO 9508568 30 Mar 1995

APPLICATION INFO:

WO 94-US10729 21 Sep 1994

PRIORITY INFO:

US 93-124771 21 Sep 1993

DOCUMENT TYPE:

Patent

LANGUAGE:

English

OTHER SOURCE:

WPI: 95-139554 [18]

AB A polypeptide is claimed which has a mol.wt. of 38,000 as determined by

2-dimensional gel electrophoresis and which is produced in infectious Borrelia burgdorferi (Bb). Also disclosed are other Bb polypeptides

(particularly 30 kDa) and nucleic acids encoding the polypeptides. More specifically, Bb is strain SH2 (5A3, 5A4 or 5A5. The polypeptide is used

for the production of vaccines and in the production of antibodies and for detection. In particular, detection of the presence of antibodies to the polypeptide in the blood of subjects can be used for detecting the

presence of Lyme disease. The 38 kDa

polypeptide is found only in highly infective Bb strains and not in low infective Bb strains and can provide selective and sensitive assays for

Lyme disease and effective vaccines. (76pp)

L61 ANSWER 20 OF 25 MEDLINE

DUPLICATE 13

ACCESSION NUMBER:

96016737

MEDLINE

DOCUMENT NUMBER:

96016737

TITLE:

Genetic and phenotypic analysis of Borrelia miyamotoi sp. nov., isolated from the ixodid tick Ixodes persulcatus, the

AUTHOR:

McClelland M; Nakao M  
Fukunaga M; Takahashi Y; Tsuruta Y; Matsushita O; Ralph D;

CORPORATE SOURCE:

Faculty of Pharmacy and Pharmaceutical Sciences, Fukuyama University, Hiroshima, Japan..

CONTRACT NUMBER:

AI34829 (NIAID)

SOURCE:

INTERNATIONAL JOURNAL OF SYSTEMATIC BACTERIOLOGY, (1995 Oct) 45 (4) 804-10.

PUB. COUNTRY:

United States

LANGUAGE:

English

FILE SEGMENT:

Priority Journals

OTHER SOURCE:

GENBANK-D45192; GENBANK-L40596; GENBANK-L40597; GENBANK-L40597; GENBANK-M64311; GENBANK-M75149; GENBANK-M60968; GENBANK-M60970; GENBANK-M64312

ENTRY MONTH:

199601

AB

The ixodid tick Ixodes persulcatus is the most important vector of Lyme disease in Japan. Most spirochete isolates obtained from I. persulcatus

ticks have been classified as Borrelia burgdorferi sensu lato because of their genetic, biological, and immunological characteristics. However, we found that a small number of isolates obtained from I. persulcatus contained a smaller 38-kDa endoflagellar protein and

single 23S-5S rRNA gene unit. Representative isolate HT31T (T = type

Searched by Barb O'Brien, STIC 308-4740



strain) had the same 23S rRNA gene physical map as *Borrelia turicatae*. The DNA base composition of strain HT31T was 28.6 mol% G+C. DNA-DNA hybridization experiments revealed that strain HT31T exhibited moderate levels of DNA relatedness (24 to 51%) with *Borrelia hermsii*, *B. turicatae*, *Borrelia parkeri*, and *Borrelia coriaceae*. However, the levels of DNA reassociation with the previously described Lyme disease borreliae (*B. burgdorferi*, *Borrelia garinii*, and *Borrelia azelii*) were only 8 to 13%. None of the previously described species examined exhibited a high level of DNA relatedness with strain HT31T. In addition, the 16S rRNA gene sequence (length, 1,368 nucleotides) of strain HT31T was determined and aligned with the 16S rRNA sequences of other *Borrelia* species. Distance matrix analyses were performed, and a phylogenetic tree was constructed. The results showed that isolate HT31T is only distantly related to both previously described Lyme disease borreliae and relapsing fever borreliae. Thus, the spirochetes isolated from *I. persulcatus* and closely related isolates should be classified as members of a new *Borrelia* species. We propose the name *Borrelia miyamotoi* sp. nov. for this spirochete; strain HT31 is the type strain.

L61 ANSWER 21 OF 25 MEDLINE

ACCESSION NUMBER: 94156457 MEDLINE

DOCUMENT NUMBER: 94156457

TITLE: Epitopes shared by unrelated antigens of *Borrelia burgdorferi*.

AUTHOR: Anda P; Backenson P B; Coleman J L; Benach J L

CORPORATE SOURCE: Centro Nacional de Microbiología, Virología e Inmunología

SOURCE: Sanitarias, Instituto de Salud Carlos III, Madrid, Spain. INFECTION AND IMMUNITY, (1994 Mar) 62 (3) 1070-8.

JOURNAL code: G07. ISSN: 0019-9567.

PUB. COUNTRY: United States

JOURNAL: Article; (JOURNAL ARTICLE)

LANGUAGE: English

FILE SEGMENT: Priority Journals; Cancer Journals

ENTRY MONTH: 199406

AB

An immunoglobulin M kappa-chain murine monoclonal antibody (CAB) reacted in a Western blot (immunoblot) with approximately 30 polypeptides from a whole-cell lysate of several American and European *Borrelia burgdorferi* strains. The reactive antigen with the highest  $M(r)$  was measured at 93 kDa (p93) and had an NH2-terminal sequence identical to the one previously reported for this antigen. The lowest reactive antigen had an  $M(r)$  of 16,000. All antigens recognized by CAB had isoelectric points within a narrow acidic range, between 5.4 and 6.2. Thus, the objective of this study was to determine whether the broad reactivity of CAB could be due to degradation of the antigen with the highest  $M(r)$ , since such spontaneous degradation of p93 has already been reported, and to determine whether CAB could recognize shared epitopes in different antigens. Treatment of *B. burgdorferi* with protease inhibitors did not result in changes in CAB reactivity, indicating that if such degradation existed, it was most likely not due to the action of endogenous proteases. Likewise, protease treatment of intact organisms and recovery of the antigens in the insoluble fraction of a Triton X-114 partition indicated that they were internal and thus less likely to be degraded by experimental procedures. Amino-terminal sequences of other reactive polypeptides showed one approximately 72-kDa polypeptide to be identical to the Dnak homolog of *B. burgdorferi*. Two other antigens at approximately 49 and 47 kDa were blocked to Edman degradation. Finally, one sequenced polypeptide with a molecular mass of approximately 38.5 kDa had a strong

identity with glyceraldehyde-3-phosphate dehydrogenase of other bacteria and vertebrates. Thus, while it cannot be ruled out that some of the CAB reactivity may be due to fragmentation of p93, there is strong evidence to indicate the presence of a shared epitope in at least three, possibly five, unrelated antigens of *B. burgdorferi*. A linear epitope within amino acid residues 357 to 371 of p93 was identified. Evidence is presented for

Searched by Barb O'Brien, STIC 308-4740





a discontinuous epitope in the carboxy-terminal region of the Dnak homolog, which bears strong amino acid identity with the p93 epitope. The conserved amino acid sequences necessary for these shared epitopes indicate possible genetic and/or functional relatedness among these various antigens.

DUPLICATE 15

L61 ANSWER 22 OF 25 MEDLINE  
 95213112 MEDLINE  
 95213112 DOCUMENT NUMBER:  
 TITLE: Ultrastructure of *Borrelia burgdorferi* after exposure to benzylpenicillin.  
 AUTHOR: Schaller M, Neubert U  
 CORPORATE SOURCE: Dermatologische Klinik, Ludwig-Maximilians-Universität, München, Germany.  
 SOURCE: INFECTION, (1994 Nov-Dec) 22 (6) 401-6.  
 Journal code: G08, ISSN: 0300-8126.  
 GERMANY: Germany, Federal Republic of  
 Journal; Article; (JOURNAL ARTICLE)  
 LANGUAGE: English  
 FILE SEGMENT: Priority Journals  
 ENTRY MONTH: 199507  
 AB The aim of this study was to investigate the morphological changes of *Borrelia burgdorferi* associated with penicillin treatment. An isolate of

*B. burgdorferi* from an erythema migrans lesion was cultivated in BSK II medium and exposed to increasing concentrations (0.0625 mg/l-2 mg/l) of penicillin G for 5 days. The in vitro minimal inhibitory concentration (MIC) was determined to be 0.5 mg/l by broth dilution method. The morphological structures of untreated spirochetes, as well as their characteristic ultrastructural changes when exposed to penicillin, were observed by electron microscopy. The following alterations were discovered: (1) Numerous **outer sheath** blebs at a penicillin concentration of 0.0625 mg/l. (ii) A characteristic irregular waveform of the borrelial cells and complete loss of the **outer sheath** at a penicillin concentration of 0.125 mg/l. (iii) The presence of "spheroplasts" at the same concentration. (iv) Structural changes of the protoplasmic cylinder complex which showed an irregular pattern at a penicillin concentration of 0.125 mg/l. (v) Disruption of the protoplasmic cylinder complex into several parts at penicillin concentrations of 0.25 mg/l and 0.5 mg/l. (vi) Severe cytolysis at penicillin concentrations of 1 mg/l and 2 mg/l.

L61 ANSWER 23 OF 25 CAPLUS COPYRIGHT 1999 ACS

ACCESSION NUMBER: 1993:249317 CAPLUS  
 DOCUMENT NUMBER: 118:249317  
 TITLE: Bacterial expression vectors containing DNA encoding secretion signals of lipoproteins and their uses for preparation of vaccines  
 INVENTOR(S): Stover, Charles K.  
 PATENT ASSIGNEE(S): Medimmune, Inc., USA  
 SOURCE: PCT Int. Appl., 115 pp.  
 CODEN: PIXXD2  
 DOCUMENT TYPE: Patent  
 LANGUAGE: English  
 FAMILY ACC. NUM. COUNT: 2  
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9307897	A1	19930429	WO 1992-US9075	19921021
W: AU, CA, JP				
RW: AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE				
AU 9229110	A1	19930521	AU 1992-29110	19921021
AU 681572	B2	19970904		

Searched by Barb O'Brien, STIC 308-4740



EP 625052 A1 19941123 EP 1992-923207 19921021  
 R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LI, LU, MC, NL, SE  
 JP 07502646 T2 19950323 JP 1992-507931 19921021  
 US 1991-780261 19911021 WO 1992-US9075 19921021

AB A bacterial expression vector contg. a DNA encoding the secretion signal of a lipoprotein and a heterologous protein antigen is prepd. The expression vector increases the immunogenicity of the protein by enabling the presentation of the protein on the surface of the bacterial host. Transformed bacteria expressing a chimeric gene for a fusion protein of the lipoprotein and the antigen protein can be used in a vaccine. Mycobacterium such as BCG may be transformed with a plasmid vector encoding an outer surface protein A or B of *Borrelia burgdorferi* and used in vaccines against Lyme disease. Construction of plasmids, e.g. p2619::Ospa, contg. the BCG HSP60 gene promoter, the signal sequence of the 19-kDa antigen of *M. tuberculosis*, and the Ospa antigen was demonstrated. Immunization of mice with the BCG transformed with plasmids was also shown.

L61 ANSWER 24 OF 25 BIOTECCHDS COPYRIGHT 1999 DERWENT INFORMATION LTD  
 ACCESSION NUMBER: 92-05816 BIOTECCHDS  
 TITLE: Specific immunofluorescent staining of pathogenic treponemes with a monoclonal antibody;  
 Treponema pallidum subsp. pallidum detection for syphilis diagnosis  
 AUTHOR: Ito F; Hunter E F; George R W; Pope V; \*Larsen S A  
 LOCATION: Division of Sexually Transmitted Diseases Laboratory  
 Research, Centers for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia 30333, USA.  
 SOURCE: J.Clin.Microbiol.; (1992) 30, 4, 831-38  
 CODEN: JCMIDW  
 DOCUMENT TYPE: Journal  
 LANGUAGE: English  
 AB BAB/c mice (4-5-wk-old) were immunized with Percoll-purified Treponema pallidum subsp. pallidum street strain DAL-1 by i.p. injection in phosphate-buffered saline-Freund's incomplete adjuvant. 3 More injections in incomplete adjuvant were given on days 15, 29 and 66, with a final booster on day 70. Spleen cells were fused with SP2/0 plasmacytoma cells by mixing in 4:1 ratio in the presence of PEG 1500 at 37 deg. Hybrids were selected in HAT medium and assayed by a microimmunofluorescent antibody and ELISA procedure. Selected cultures were cloned twice by limiting dilution. Monoclonal antibodies (MAbs) T10/2 and T1P/73 reacted with the 37 kDa molecule of DAL-1, and with the 37 kDa antigen of T. pallidum subsp. pallidum Nichols and T. pallidum subsp. pertenue Gauthier, but did not react with Treponema phagedenis Reiter, Treponema denticola MRB, Treponema reitingeri Noguchi, or other spirochetes, such as *Borrelia burgdorferi* and Leptospira interrogans serovar pomona, or with normal rabbit testicle tissue. The MAbs were labeled with fluorescein isothiocyanate and used to diagnose syphilis by direct staining of lesion exudate or formalin-fixed tissue. (38 ref)

L61 ANSWER 25 OF 25 CABA COPYRIGHT 1999 CABI  
 ACCESSION NUMBER: 92:70811 CABA  
 DOCUMENT NUMBER: 922268208  
 TITLE: Serodiagnostics of leptospirosis in pigs using an axial filament enzyme-linked immunosorbent assay  
 AUTHOR: Mendoza, L.; Prescott, J. F.  
 CORPORATE SOURCE: Department of Veterinary Microbiology and Immunology, University of Guelph, Guelph, Ont. N1G 2W1, Canada.  
 SOURCE: Veterinary Microbiology, (1992) Vol. 31, No. 1, pp. 55-70. 32 ref.

Searched by Barb O'Brien, STIC 308-4740



The axial filament (AF) from *Leptospira interrogans* serovar canicola was isolated by caesium chloride density gradient centrifugation of 2% sarcosyl treated whole cells. Isolation of AF was confirmed by electron microscopic examination, by protein-A immunogold labelling, sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), and immunoblotting. Analysis by SDS-PAGE of the purified preparation showed relatively weak bands of molecular size 42 kDa and 21 kDa and strong bands of 35 kDa and 34.5 kDa. Immunoblot analysis using antiserum to the AF against sonicated leptospirae of a variety of serovars showed prominent reaction against the 41, 35 and 34.5 kDa protein bands, as well as against minor bands of molecular weight 43, 39 and 37 kDa. Antisera prepared against leptospiral serovars also identified minor bands at 33 and 32 kDa. Immunoblots with antiserum to whole cells of serovar Bratislava detected the 35 and 34.5 kDa AF bands to *Borrelia burgdorferi* moderately and of *Treponema hyodysenteriae* only slightly in comparison to leptospiral AF. Antibody of *B. burgdorferi* did not detect the leptospiral AF antigen. Immunoblots with antiserum to *T. hyodysenteriae* showed a marked reaction with a 41 kDa band of *B. burgdorferi* but only a very minor reaction with leptospiral AF. The AF was tested in an AF-ELISA against sera from 260 pigs, many of which reacted in the microscopic agglutination test (MAT) against one or more leptospiral serovars. A sensitivity of 97.1% and specificity of 93.1% was determined in comparison to the MAT ( $r = 0.4$ ). Only moderate correlation was observed between titres detected in the AF-ELISA and the MAT. When sonicated whole cells (WC) of serovar canicola were used in an ELISA (WC-ELISA), high correlation was observed between AF-ELISA and WC-ELISA ( $r = 0.97$ ). It was concluded that the AF-ELISA can be used effectively as a species-specific antigen for the serological diagnosis of leptospirosis in swine and that sonicated whole cells can substitute excellently for purified AF as the antigen source. It is suggested that these findings may be extrapolated to the use of AF in immunodiagnosis of leptospirosis in other species.

FILE 'HOME' ENTERED AT 14:24:36 ON 19 AUG 1999

